Move over, LDL. Ceramides may be better markers of heart risk p. 1080

Colorectal cancer rises in younger people p. 1088

Protein structure prediction with language models p. 1123

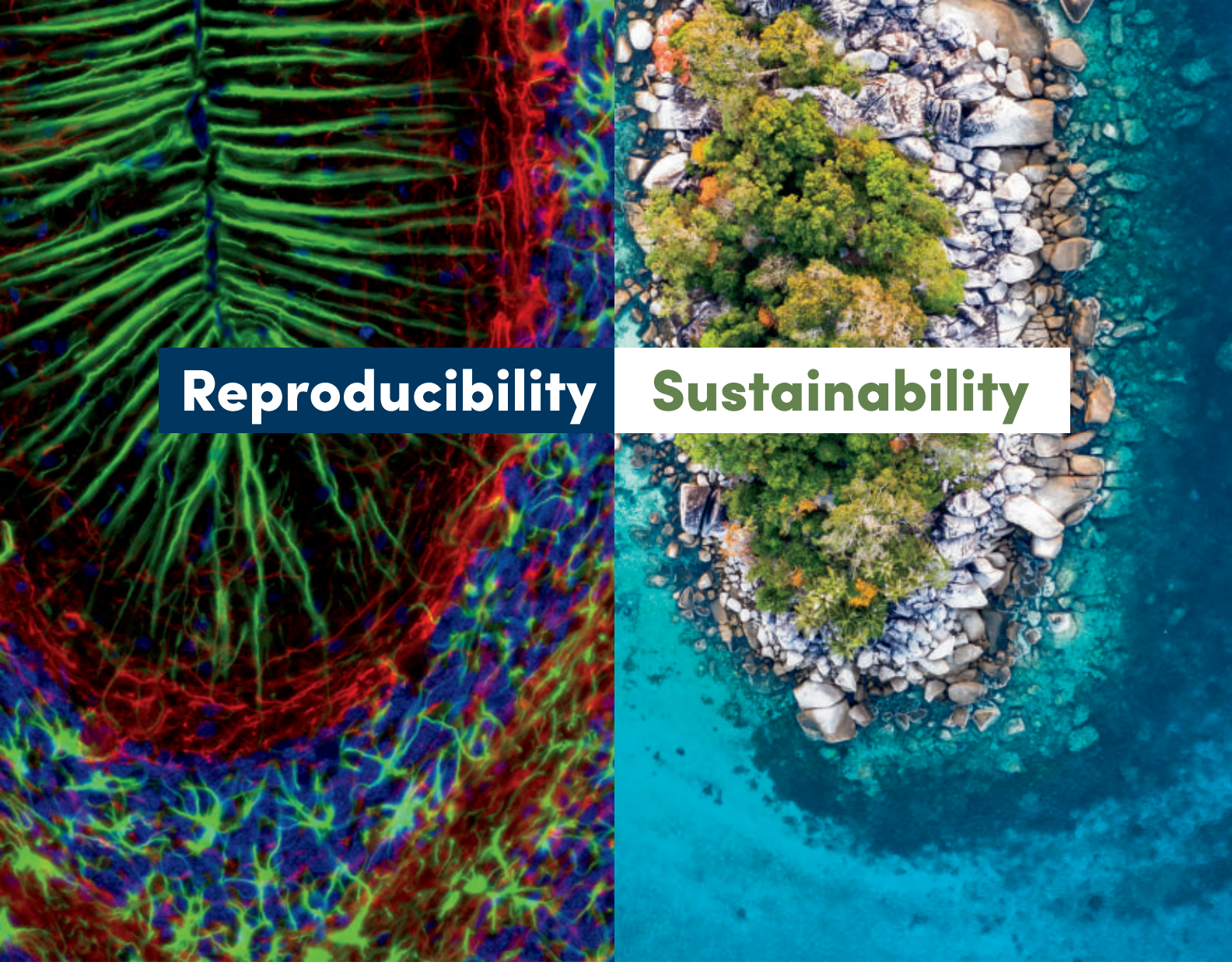
Science

$15
17 MARCH 2023
science.org

AAAS

LOOK OUT

Evolving warning coloration p. 1136



CST scientists have always been dedicated to helping researchers find answers and identify solutions. As scientists, we've always been passionate about sustainability, too. And now more than ever, sustainable solutions are needed. CST has committed to achieving net-zero emissions by 2029. And we've joined 1% for the Planet as its first life sciences industry member. Doing good science is important. So is just doing good.

www.cellsignal.com

Cell Signaling TECHNOLOGY®

1% FOR THE PLANET

Antibodies for Research

Immunoassay Kits

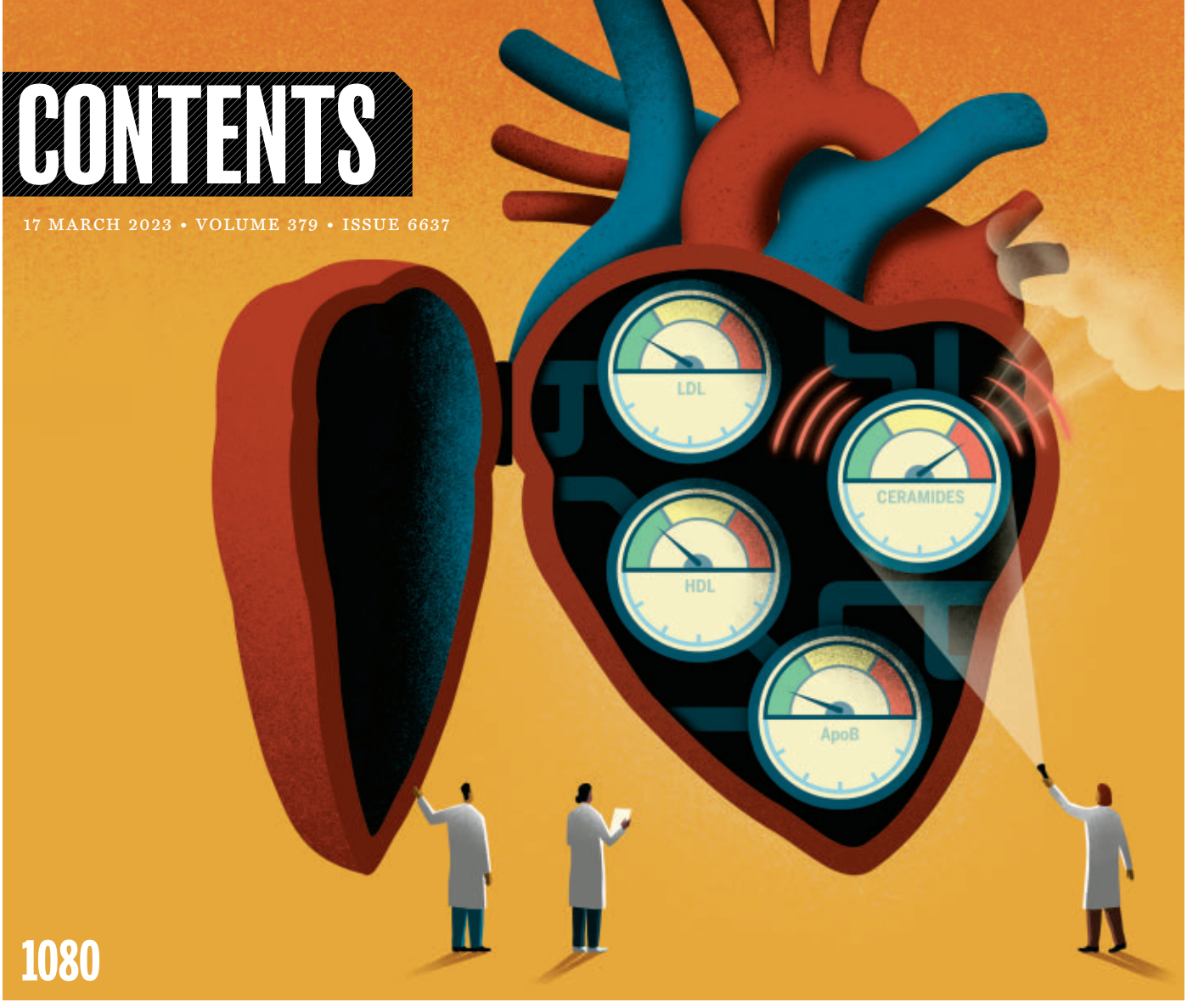
Customs & Services

Application Workflow Solutions

Industry-leading Validation Process

CONTENTS

17 MARCH 2023 • VOLUME 379 • ISSUE 6637



1080

NEWS

IN BRIEF

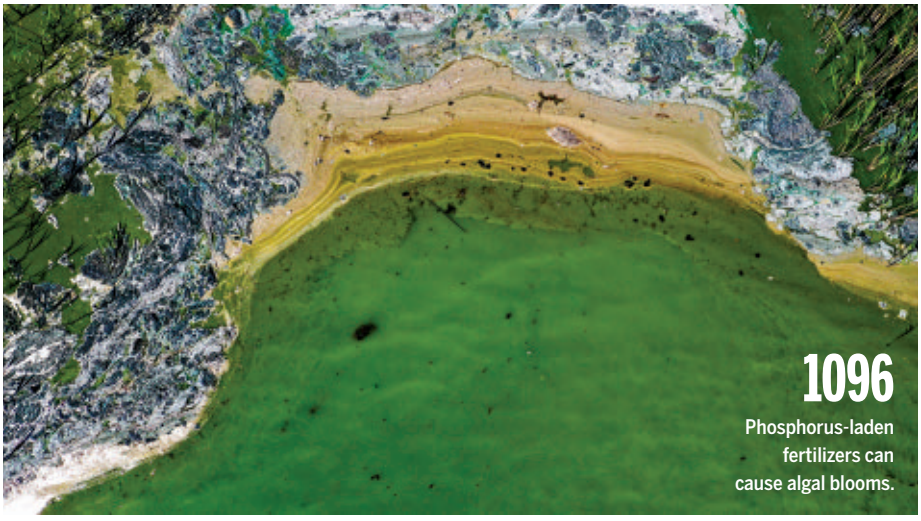
IN DEPTH

FEATURES

INSIGHTS

POLICY FORUM

PERSPECTIVES



1096
Phosphorus-laden fertilizers can cause algal blooms.

RESEARCH

ON THE COVER

An alpine newt (*Ichthyosaura alpestris*) walks to a breeding pond in the Alps, France. Many amphibians have a cryptic upper side but a normally concealed, conspicuous underside. These hidden signals have evolved for several reasons, including as a warning display to would-be predators. A phylogenetic analysis reveals that species with hidden colors represent an important intermediate step in the evolution of species with permanently displayed warning signals.
See page 1136.
Photo: Cyril Ruoso/NPL/Minden Pictures

PHOTO: PONTUS LUNDAHL/GETTY IMAGES

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): $2411; Foreign postage extra: Air assist delivery: $107. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

EDITORIAL

Let's change what's possible

America's science, technology, and innovation ecosystem is a powerful engine for progress, but it was conceived in the last century for last century's goals. Today, the nation's aspirations have never been bigger: robust health and ample opportunity for everyone, tackling the climate crisis and using it to reimagine infrastructure and humanity's relationship with nature, global security and stability, a competitive economy that creates good-paying jobs, and a strong, thriving democracy. The purpose of science and technology is to open the doors that make these aspirations possible. As President Biden said, "We can channel the full talents of all our people into a greater measure of hope and opportunity for our nation and for the world."

Building that equitable, resilient, and ambitious future starts with federal research and development (R&D). And the magnitude of today's challenges means that it's time for purposeful steps that boost the federal R&D enterprise to meet our aspirations. One step is to continue current federal R&D investments, including those in basic research. It's time to renew the vibrancy of research, open participation to a more diverse community of people and institutions, and recommit to the many national purposes behind public R&D spending, including improving health outcomes, creating more economic opportunity and industries of the future, and strengthening national security. That's exactly why President Bident has proposed historic federal investments in R&D, including $210 billion in his latest budget released last week.

> **"...it's time for purposeful steps that boost the federal R&D enterprise to meet our aspirations."**

Another step is addressing the gap between the country's excellent research and the societal impact we seek: tangible benefits for people in every community to live better lives. To get there, research must translate into new products and services, new industries and jobs, new policies and regulations, and new standards and practices. In this vein, the National Science Foundation's Directorate for Technology, Innovation and Partnerships is helping universities move basic research into commercialization and is boosting regional innovation. And under the CHIPS and Science Act, the National Institute of Standards and Technology is building frontier semiconductor R&D to reinvigorate a critical domestic industry.

Additionally, some investments must take aim at bold, barely feasible goals. One example is President Biden's Cancer Moonshot, which aims to reduce the age-adjusted cancer death rate by at least 50% in 25 years and to improve the experience of patients, families, and caregivers who are dealing with cancer. These challenging goals are mobilizing people and organizations across government and the private sector in ways that will change millions of lives. A complementary example is the adaptation of the Defense Advanced Research Projects Agency (DARPA) model to build capacity for developing breakthroughs in other sectors. DARPA's "What does it take?" mentality pulls innovators together to build on each other's work, take risk, fail, and try again until a seemingly impossible goal is achieved. That's the spirit that President Biden invoked when he launched the Advanced Research Projects Agency for Health (ARPA-H) last year.

As well, the power of R&D must be brought to missions that have not historically been the focus of innovation. Almost all federal R&D is aimed at national security, space, health, energy, the environment, and the country's basic research foundation. But today's research and technological advances can create possibilities for a much wider array of needs—from K-12 education to workforce training to construction to traffic safety. Efforts in the Department of Education and the Department of Transportation are now exploring new R&D investments that can achieve better outcomes for their missions.

These are important shifts, and they invite every member of the R&D community to step up to new challenges. For early-career scientists and engineers, this is an invitation to imagine the future you want to live in and to find or create ways to pursue bold R&D. For managers and leaders, this is an invitation to lift your teams up by imbuing them with a passion for purpose. And for every person who seeks to innovate, this is an invitation to bring your personal perspective—reflecting who you are and where you come from—to help shape a future in which every person can thrive.

–Arati Prabhakar

PHOTO: CHRISTOPHER MICHEL

Arati Prabhakar is the director of the White House Office of Science and Technology Policy, assistant to the president for Science and Technology, and a member of President Biden's Cabinet, Washington, DC, USA. press@ostp.eop.gov

10.1126/science.adh5328

15.1 Length, in meters, of the neck of a late Jurassic sauropod dinosaur, *Mamenchisaurus sinocanadorum*, estimated from an incomplete specimen discovered in China in 1987. That neck span—longer than a typical school bus—exceeds the neck of any other known sauropod. (*Journal of Systematic Palaeontology*)

IN BRIEF

Edited by **Jeffrey Brainard**

OVERSIGHT

Congress pursues COVID-19 origin

Researchers and lawmakers are waiting to see whether President Joe Biden will sign a bill finalized by Congress last week that would declassify more information from U.S. intelligence agencies about the origin of the COVID-19 pandemic. Biden's administration previously said four U.S. intelligence agencies lean toward a natural origin of the pandemic—a view shared by most outside scientific experts—whereas two favor a laboratory-leak explanation. Two others are undecided. The bill sent to Biden, approved unanimously by both chambers of Congress, asks the U.S. director of national intelligence (DNI) to "declassify any and all information" on potential links between the pandemic and the Wuhan Institute of Virology in China. It would give the DNI 90 days to comply, while allowing the director to hold back information that would compromise intelligence sources. Even if the measure becomes law, some doubt it will reveal anything that will settle the contentious origin question.

China reworks R&D management

POLICY | The Chinese government has unveiled a major shake-up of its science and technology bureaucracy, aiming for "self-reliance" in R&D. A restructuring plan approved on 13 March by the National People's Congress creates a new, high-level Central Science and Technology Commission that will give the Communist Party greater control of broad R&D strategy while streamlining the role of the Ministry of Science and Technology (MOST). Direct oversight of research deemed less critical to China's development and global competitiveness will move from MOST to other agencies. The agriculture ministry, for example, will take over farm science. Many details have yet to be released, but the plan represents "the most radical change to [China's] innovation system since the end of the Mao era," says Richard Suttmeier, a political scientist retired from the University of Oregon.

U.S. research reactor can restart

NEUTRON RESEARCH | This week, regulators gave the go-ahead to restart a small U.S. research reactor, 2 years after an accident shut it down and cut almost by half the nation's capacity to use neutrons to study materials. The February 2021 incident at the 20-megawatt, neutron-generating reactor at the National Institute of Standards and Technology (NIST) in Gaithersburg, Maryland, occurred when one of its 30 rodlike, uranium-filled fuel pins popped out of place, overheated, and partially melted. A trace amount of radiation escaped the building but did not jeopardize the public, NIST said. In August 2022, the U.S. Nuclear Regulatory Commission (NRC) and NIST agreed to procedural changes that would ensure the pins are always correctly latched. NRC now reports it is satisfied with NIST's corrective actions. The reactor will power up slowly over several weeks.

Removing race from genetics

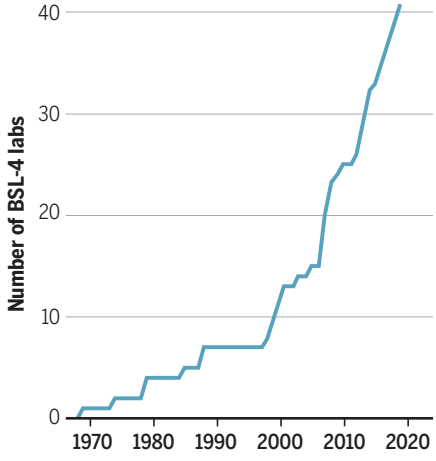
DIVERSITY | Genomics researchers should not use race to describe a population's genetic ancestry and instead should use terms carefully tailored for accuracy, a U.S. national academies panel said this week. The National Academies of Sciences, Engineering, and Medicine released the report after the National Institutes of Health requested information on how to describe populations in genomics studies. The panel's report concludes that the notion that people belong to genetically distinct races is scientifically invalid. And it recommends using people's ethnicity, such as Latino; geographic location, such as Japanese; or region of ancestry, such as African, only in certain cases. Researchers studying disparities in health care may want to use racial categories because they can serve as a proxy for people's experience of structural racism in health settings, the panel said. Studies of disease genes or human evolution should describe populations mainly using "genetic similarity," or how closely members' genes are related to reference genomes drawn from certain populations, such as the Yoruba of Nigeria or Tuscans in Italy.

High-risk pathogen labs mushroom

BIOSAFETY | The number of labs with the containment features needed to study

A spurt in biosafety labs

Europe has the most biosafety level-4 (BSL-4) labs, and three-quarters are in urban areas. (Ten existing labs without known start dates are not shown.)



CREDITS: (GRAPHIC) D. AN-PHAM/*SCIENCE*; (DATA) *GLOBAL BIOLABS REPORT 2023*. HTTPS://WWW.GLOBALBIOLABS.ORG

Floodwaters from an atmospheric river engulfed train cars, vehicles, and homes last week in Pajaro, California, after a levee broke.

PHOTO: JOSH EDELSON/AFP/GETTY IMAGES

METEOROLOGY

Intensity scale for atmospheric rivers reveals global hot spots

Atmospheric rivers like those pummeling the West Coast now have a five-level intensity scale, which has enabled researchers to chart the global prevalence of these sinuous bands of storms. The scale, first developed in 2019 for the U.S. West Coast, classifies the events based on how long they last and how much moisture they transport from the tropics to higher latitudes, much as the hurricane scale ranks storms by their wind speeds. In a new study using 40 years of observations, scientists found the most extreme rivers—AR-5 on their scale—occur once every 2 or 3 years and are less likely to make landfall than weaker ones. But when the storms do hit, they tend not to penetrate inland and end up dumping all their moisture in coastal areas, causing major floods—like the AR-5 storm that hit Pakistan last year and the AR-4 storm that struck California in January. Hot spots for AR-5s include British Columbia and the Pacific Northwest, northwestern Europe, and southern Chile. The research was published last month in the *Journal of Geophysical Research: Atmospheres*.

the most deadly known human pathogens is booming, raising risks of an accidental release or use by a terrorist, warns an analysis. Fifty-one biosafety level-4 (BSL-4) labs exist worldwide, roughly double the number a decade ago, says the *Global BioLabs Report 2023*, published this week by researchers at King's College London and George Mason University. The report also documented 57 BSL-3 "plus" labs, many of which study animal pathogens. BSL-4 labs enable studies of threats such as the viruses that cause Ebola, and 18 more of the facilities are slated to open in coming years—most in Asian countries such as India and the Philippines that want them to bolster responses to local threats and future pandemics. But many countries lack strong methods to monitor such labs, the report's authors say. They urge the World Health Organization to strengthen lab safety guidance and want individual countries to agree to outside audits to ensure their BSL-4 labs meet international standards.

Maternal, child death rates jump

PUBLIC HEALTH | Mortality rates for U.S. children, teenagers, and pregnant people grew strikingly in 2021, according to data published this week. The maternal mortality rate has nearly doubled since 2018, and increases in 2020 and 2021 for children and teens were the largest in decades. Maternal mortality is defined as deaths from pregnancy-related causes during pregnancy or within 6 weeks after. It grew from 23.8 deaths per 1000 live births in 2020 to 32.9 deaths per 1000 in 2021, the U.S. Centers for Disease Control and Prevention (CDC) said. The rate among Black mothers, 69.9 deaths per 1000, was 2.6 times that of non-Hispanic white mothers. CDC did not give a reason for the growth, but other reports have cited COVID-19. Among children and teens ages 1 to 19, the mortality rate jumped by 8.3% in 2021 after growing by 10.7% in 2020, *The Journal of the American Medical Association* reported. The increase was driven by murders, suicides, and fatalities associated with traffic accidents and drug overdoses.

Dance about materials wins prize

SCIENCE COMMUNICATION | Twirling hand fans, catchy *Lord of the Rings* references, and 20 blue papier-mâché balloons helped illustrate a video about crystalline materials (metal-organic frameworks) that won *Science*'s long-running Dance Your Ph.D. contest this year. The video, put together by University of Oregon chemist Checkers Marshall, aimed to explain their thesis on the frameworks, which are made up of metal ions bound to molecules. Because of the materials' porous nature, the frameworks can act like sponges and capture gases such as carbon dioxide. The dance video, in which the blue balloons stood in for ions, depicted how Marshall's Ph.D. work aims to make these materials more effective for other applications, such as water filtration and nerve agent detoxification. Now in its 15th year, the contest received 28 entries from 12 countries. The overall champion receives $2500. Marshall's video and other entries can be seen at https://scim.ag/dancePhDwinner.

The University of Michigan will require the latest COVID-19 vaccine booster for students living on campus in the fall.

PHOTO: DOMINICK SOKOTOFF/SIPA VIA AP IMAGES

SCIENCE POLICY

Do COVID-19 vaccine mandates still make sense?

Ineffective or outdated requirements could undermine trust, some vaccine researchers say

By **Gretchen Vogel**

Visitors to the National Academy of Sciences (NAS) in Washington, D.C., receive a clear reminder that, 3 years after the World Health Organization (WHO) declared COVID-19 a pandemic on 10 March 2020, it's far from over. Before entering, they must show a guard proof that they have been vaccinated against COVID-19. Such demands were common around the world a year ago, with wide support from infectious disease scientists and public health researchers. But by now, almost everyone has had natural infections with SARS-CoV-2 or been vaccinated against the coronavirus—sometimes both—and it's become clear that vaccine-induced immunity quickly loses its ability to prevent infection and spread of the latest variants. Some now say the mandates are outdated.

The persistent requirements are "baffling to say the least," says Heidi Larson, an anthropologist at the London School of Hygiene & Tropical Medicine and director of the Vaccine Confidence Project. She spoke at a major infectious disease meeting this year that required all attendees to show they had had two doses of a vaccine—with no need for a recent booster. "It's not like it's going to mitigate the spread."

Larson and other vaccine acceptance researchers who spoke to *Science* all emphasize that COVID-19 vaccines clearly prevent severe disease, but they worry maintaining the mandates could undermine future public health efforts. "Having to show these old vaccination proofs or certificates really doesn't make sense, and it could cause harm, because people might lose trust in the competence of the organization," says University of Konstanz psychologist Katrin Schmelz, whose research has found that institutional trust is crucial for health policy acceptance.

Mandates became common in 2021 and early 2022, after the Delta variant caused new peaks of COVID-19 hospitalizations and deaths, especially among people who had not been vaccinated. Across Europe, people had to show they were fully vaccinated before entering restaurants, shops, museums, and concert halls. The United States required federal employees to be fully vaccinated to keep their jobs. Singapore imposed a similar mandate on all employees, both public and private. And in February 2022, after months of debate, Austria passed one of the world's first nationwide vaccine mandates, requiring the shots for all residents over age 18 and imposing fines on those who refused.

In many places, the mandates sparked large protests, but the justifications seemed compelling. COVID-19 vaccines offer powerful protection against severe disease, so the measures promised to keep hospitals from being overwhelmed. Early data also suggested the vaccines reduced overall infections and shortened the time a person was infectious. "If you can only transmit for 3 days, that's much better than 7 days," says Angela Branche, an infectious disease expert who studies vaccine efficacy at the University of Rochester.

Initial hopes that the vaccines would stop the spread of COVID-19 faded, however, as it became clear that protection against infection wanes after a few months. New variants that could get around vaccine-induced immunity further undermined hopes that the shots could curb spread.

In April 2022, researchers in the United Kingdom reported in *The New England Journal of Medicine* that, based on the health records of more than 1.5 million people, protection against symptomatic COVID-19 with the Omicron variant faded to zero 25 weeks after a second shot of the AstraZeneca vaccine and to just 9% 25 weeks after a second dose of the Pfizer-BioNTech vaccine. A booster dose increased protection back up above 60% for a month or two, but by 10 weeks that protection had also started to wane. (Protection against severe disease persists longer.) Now that ever-larger numbers of people have some immunity after natural infections, the real-world benefits of vac-

cines have become still harder to measure.

Many places and groups soon lifted their vaccine requirements or stopped enforcing them. In June, Austria revoked its law. Most European countries that required vaccine "passes" for shopping, eating out, and more dropped them by summer 2022. In October, Singapore announced it was lifting its vaccine mandate, and 1 month later the German health minister announced that even for health care workers, its vaccine requirement would be allowed to expire. Because being vaccinated was no longer a significant protection against infections with the newer variants, he said, "there's no longer a reason," epidemiologically, for the mandate.

Compared with Europe and Asia, the United States appears to be holding on to vaccine mandates more tightly. Many U.S. scientific groups, including NAS and AAAS (publisher of *Science*) still require their employees and all attendees at events and meetings to be vaccinated. Many universities continue to require vaccination or booster shots for students, staff, or both.

Although the U.S. government stopped enforcing a federal employee mandate last year in the face of lawsuits, it retains other requirements. Foreign citizens entering the country still need to show they have received a course of WHO-approved shots, a requirement that made the news last month when tennis star Novak Djokovic, who is not vaccinated, requested an exemption to compete in March tournaments in Florida. (His request was denied.)

Scientists traveling to some meetings face similar requirements. Those attending the American Astronomical Society annual gathering in January had to upload proof of vaccination, including one booster, before registering for the meeting. Larson and other attendees at the Conference on Retroviruses and Opportunistic Infections, held in February, had to show they had received two doses of a vaccine. At AAAS's annual meeting this month, in-person attendees were also required to confirm they were vaccinated, albeit on an honor system.

None of those meetings specified the vaccination had to be recent—so attendees at some of the gatherings may have gotten their last shot more than 18 months ago. Nor did the meetings accept evidence of an infection with SARS-CoV-2, recent or otherwise, as an alternative. That doesn't make sense to Maxwell Smith, who studies public health ethics at Western University. "If they say you need to have been vaccinated, but nothing about when those vaccines were received, nor anything about recent infection, then of course that's less likely to achieve the objectives" of reducing transmission and infections, he says. "It would be more justifiable to say you need to have received a vaccine or been infected in the past 3 to 6 months, for example."

Political scientist Katie Attwell, who studies vaccine policy and acceptance at the University of Western Australia, Perth, agrees. Asking for just two doses at some time in the past "feels strange and out of date," she says. "If it was a living policy, you'd mandate the boosters."

Branche echoes the worry that many of the lingering mandates could be counterproductive. "We don't want people to think they're safe from getting infected or transmitting the virus because they had the primary series 2 years ago. That's just not true," she says, adding that such policies might also discourage people from getting further shots.

Others say the conference vaccine requirements may be substituting for more effective ways to prevent spread of COVID-19. "If I were to see a meeting that had a vaccine requirement but then put everyone in the standard ballroom chairs shoulder to shoulder without a mask requirement, I might not consider that meeting seriously focused on COVID protection," says University of Maryland School of Medicine epidemiologist Meagan Fitzpatrick, who models infectious disease transmission. "The vaccine requirement does not make it OK to drop all the other efforts that one might be able to deploy."

> **"If it was a living policy you would mandate the boosters."**
> **Katie Attwell**, University of Western Australia, Perth

Many organizations are reviewing or revising their vaccine policies, especially as the end of the U.S. COVID-19 emergency declaration, set for 11 May, draws near. NAS, for example, told *Science* it is reassessing its current mandate. The University of Michigan, which had required all students, faculty, and staff to be vaccinated and receive a booster dose, announced in February that only students living in on-campus housing will be subject to a mandate. They will be required to have a dose of the bivalent booster, available since September 2022 and designed to protect against the original strain of SARS-CoV-2 as well as Omicron.

Rob Ernst, the university's chief health officer, says requiring the bivalent booster means that at the start of the fall semester all residents will have had a booster that is less than 1 year old. And the rule is still needed, he argues. With as many as 1200 students living in some residence halls, "the potential for disruption is greatest in that area." After 3 years, Ernst says, "We still have significant COVID in our community." ■

BIOMEDICINE

Shadowed by past, gene-editing summit looks to future

London meeting touts sickle cell success, but questions about embryo editing linger

By **Kai Kupferschmidt**

After decades of living with often excruciating pain, Victoria Gray had to get used to a new sensation in recent years: waking up without it. "It may sound crazy, but I had to pinch myself to see was I still able to feel pain," she says.

Gray, a 37-year-old mother of four from Forest, Mississippi, who was born with sickle cell disease, arguably became the star of last week's International Summit on Human Genome Editing in London when she spoke about her transformation. In 2019, she was the first person to undergo an experimental therapy in which blood stem cells were taken from her, altered with the gene editor CRISPR to compensate for the sickle cell mutation, and returned to her body.

She now produces few of the abnormally rigid, sickle-shaped red blood cells that can block blood flow, causing intense pain. "At one point in my life, I stopped planning for the future because I felt I didn't have one," Gray told a rapt audience at the summit. "Now, I can dream again without limitation."

Gray's appearance was designed to underscore the rapid clinical advances in somatic cell gene editing—making noninheritable changes to a person's DNA—and redirect attention away from the controversial prospect of heritable changes. "The Organising Committee had a clear intent … to shift the focus away from heritable to somatic," says bioethicist Françoise Baylis, now retired from Dalhousie University, who was part of that committee. That is a stark contrast to the 2018 summit in Hong Kong, which was dominated by news a few days earlier that Chinese researcher He Jiankui had used CRISPR to modify the genomes of human embryos and implanted some of them in a woman. The twin girls she gave birth to likely carry the genetic changes in their eggs and

could pass them on to subsequent generations; the announcement led to outrage, and He's imprisonment.

Since then, there has been no known attempt to produce gene-edited babies, but somatic gene-editing therapies using CRISPR or related methods have surged. Clinical trials are underway for blood disorders, cancers, diabetes, blindness, and more. The CRISPR method used to treat Gray has already been tested in more than 30 people with sickle cell disease and in more than 40 with another blood disorder, and could get U.S. approval this year.

The progress has raised a new ethical issue: how to ensure these new therapies reach those who most need them. "We're already seeing significant challenges with access to [existing] gene therapies," says Claire Booth, a gene therapy researcher at University College London. "With so much growth in this area of therapeutic gene editing the problem of delivering therapies to patients is only going to grow."

The two companies developing Gray's therapy, Vertex Pharmaceuticals and CRISPR Therapeutics, haven't set a price on it. But the procedure is complex, requiring each patient to undergo costly and somewhat risky chemotherapy to wipe out their current blood stem cells in the bone marrow to make room for cells altered outside the body. The price tag could be more than $1 million per person.

Yet over half of the more than 300,000 people born annually with sickle cell disease live in three countries where few would be able to afford that: Nigeria, the Democratic Republic of the Congo, and India. It's unclear how even the U.S. health care system can manage such costs. "It's heartbreaking because I do still have family members who suffer from sickle cell disease," says Gray, who was treated for free as part of a clinical trial. "How is it fair that something lifesaving comes with a price tag that high?"

There are other barriers. High-income countries have hundreds of bone marrow transplant centers, where stem cells can be wiped out and replaced, but sub-Saharan Africa has just three. Few countries in the region screen for the disease and give affected newborns penicillin as prophylaxis against the infections to which their condition predisposes them, says Ambroise Wonkam, a geneticist at Johns Hopkins University. "If even the simple newborn screening plus penicillin is not implemented in Africa, why should we be talking about genome editing? But I do think one does not exclude the other."

"I can dream again," says Victoria Gray, who spoke at a gene-editing summit about the treatment of her sickle cell disease.

Genetically editing cells in the body rather than outside it could cut costs. Clinical trials using such an "in vivo" approach have started for some conditions and the Bill & Melinda Gates Foundation and the National Institutes of Health have put $200 million toward developing this for sickle cell disease and HIV as well. The components of CRISPR, for example, could be loaded in delivery vehicles such as harmless viruses or lipid nanoparticles like those used to ferry in RNA for certain COVID-19 vaccines.

New ways to pay for these therapies may be needed, too. "I'm afraid that we're not ready," Steve Pearson, who heads the Institute for Clinical and Economic Review, told the summit. "I don't know how we're going to be able to create the pricing, the payment, and the intellectual property innovation at the speed that the science is bringing these treatments forward."

Bringing African scientists into the research is also an essential part of equity, says Jantina de Vries, a bioethicist at the University of Cape Town. The Gates in vivo–editing push largely funds U.S. researchers, she notes. "What happens then is that Africa is cast merely as a recipient of innovation, not a driver."

Despite the focus on somatic gene editing, the summit, the last in a series of three, could not avoid the shadow of He's experiment. Safety concerns around altering the DNA of embryos have only grown, says University of Oxford geneticist Dagan Wells.

Standard CRISPR works by breaking DNA's double-stranded helix and exploiting a cell's DNA repair machinery to insert a replacement genetic sequence. But studies have shown DNA repair is deficient in early embryos. In human embryos gene edited at fertilization, Wells found that 40% of the double-strand DNA breaks introduced by CRISPR were not repaired by the cell. If the embryos had developed, they might have suffered from severe genetic diseases.

In vitro–fertilized eggs are appealing targets for editing because the researchers have to target just one easily accessible cell. "But it also seems that that may be exactly the wrong time to do it," Wells says. "I think everything we've learned in the last few years has really cemented those kinds of concerns that were expressed back in 2018."

Newer methods such as base editing and prime editing that avoid double-break strands may be safer. "Perhaps, it'll be possible to revisit the application of genome editing to embryos with those methods," Wells says. But he cautions that their safety is unknown. "Hopefully, people will learn the lessons of the past, and they won't hurry too fast."

Some attendees felt the summit provided too little opportunity to discuss the ethical implications of embryo editing. "My suspicion is that this summit was designed to be uncontroversial, to even at points almost be boring," says Arizona State University, Tempe, bioethicist Ben Hurlbut. That way no progress is made around the real challenges in the field, he says. "And if we're not here to make progress on those things, what are we here for?"

Some saw a subtle shift in thinking on germline editing between the Hong Kong and London summits. Five years ago, summit organizers concluded it was unacceptable but added that "it is time to define a rigorous, responsible translational pathway" toward clinical trials of germline editing. This year's closing statement similarly said that "heritable human genome editing remains unacceptable at this time." But organizers also wrote that public debate must continue in order to resolve "whether this technology should be used." Baylis says that new wording is important. "It's a bit of a pulling back to acknowledge that those debates and discussions have not been concluded." ■

PHOTO: SYRIOL JONES/ROYAL SOCIETY

CHEMISTRY

Seawater splitting could help green hydrogen grow

Corrosion-proof electrolyzers could tap ample supplies of saltwater

By **Robert F. Service**

Few climate solutions come without downsides. "Green" hydrogen, made by using renewable energy to split water molecules, could power heavy vehicles and decarbonize industries such as steelmaking without spewing a whiff of carbon dioxide. But because the water-splitting machines, or electrolyzers, are designed to work with pure water, scaling up green hydrogen could exacerbate global freshwater shortages. Now, several research teams are reporting advances in producing hydrogen directly from seawater, which could become an inexhaustible source of green hydrogen.

"This is the direction for the future," says Zhifeng Ren, a physicist at the University of Houston (UH). However, Md Kibria, a materials chemist at the University of Calgary, says for now there's a cheaper solution: feeding seawater into desalination setups that can remove the salt before the water flows to conventional electrolyzers.

Today, nearly all hydrogen is made by breaking apart methane, burning fossil fuels to generate the needed heat and pressure. Both steps release carbon dioxide. Green hydrogen could replace this dirty hydrogen, but at the moment it costs more than twice as much, roughly $5 per kilogram. That's partly due to the high cost of electrolyzers, which rely on catalysts made from precious metals. The U.S. Department of Energy recently launched a decadelong effort to improve electrolyzers and bring the cost of green hydrogen down to $1 per kilogram.

If they succeed and green hydrogen production skyrockets, pressure could build on the world's freshwater supplies. Generating 1 kilogram of hydrogen using electrolysis takes some 10 kilograms of water. Running trucks and key industries on green hydrogen could require roughly 25 billion cubic meters of fresh water a year, equivalent to the water consumption of a country with 62 million people, according to the International Renewable Energy Agency.

Seawater is nearly limitless, but splitting it comes with its own problems. Electrolyzers are built much like batteries, with a pair of electrodes surrounded by a watery electrolyte. In one design, catalysts at the cathode split water molecules into hydrogen (H+) and hydroxyl (OH-) ions. Excess electrons at the cathode stitch pairs of hydrogen ions into hydrogen gas (H_2), which bubbles out of the water. The OH- ions, meanwhile, travel through a membrane between the electrodes to reach the anode, where catalysts knit the oxygen into oxygen gas (O_2) that is released.

When seawater is used, however, the same electrical jolt that generates O_2 at the anode also converts the chloride ions in saltwater into highly corrosive chlorine gas, which eats away at the electrodes and catalysts. This typically causes electrolyzers to fail in just hours when they can normally operate for years.

A green hydrogen plant in Spain will consume vast amounts of fresh water.

Now, three groups are reporting efforts to halt this corrosion. Researchers led by Nasir Mahmood, a materials scientist at RMIT University, Melbourne, reported in the 8 February issue of *Small* that by coating their electrodes with negatively charged compounds such as sulfates and phosphates, they could repel negatively charged chloride ions and prevent the formation of chlorine gas. The RMIT team reported virtually no degradation in its electrodes for up to 2 months, although it generated only a trickle of hydrogen. Since then, in unpublished work, the researchers have bolstered their setup to produce hydrogen as fast as commercial freshwater electrolyzers, Mahmood says.

Shizhang Qiao, a nanotechnologist at the University of Adelaide, and his colleagues made changes to a second type of electrolyzer that uses a membrane permeable only to H+ ions. This setup split water molecules at the anode instead of the cathode, snatching away electrons to free H+ ions. The ions migrate through the membrane to the cathode where they combine with electrons to make H_2. Qiao and his colleagues coated their electrodes with chromium oxide, which attracted a bubble of OH- ions that repelled chloride ions. The device split seawater for 100 hours at high currents without degradation, they report in the 30 January issue of *Nature Energy*. "I'm very happy to see such a clever design," says UH materials physicist Shou Chen.

Zongping Shao, a chemical engineer at the Nanjing University of Technology, and his colleagues took a third tack to fending off chloride. They surrounded the electrodes with membranes that only allow freshwater vapor to pass through from the surrounding bath of seawater. As the electrolyzer converts fresh water to hydrogen and oxygen, it creates a pressure that draws more water molecules through the membrane, replenishing the freshwater supply. In the 30 November 2022 issue of *Nature*, Shao and his colleagues reported their setup operated for 3200 hours with no sign of degradation. "It's like an internal distillation process," says Haotian Wang, an applied physicist at Rice University.

The membranes that screen out the salt resemble those in commercial desalination plants, which are already efficient enough to produce fresh water while adding only about $0.01 per kilogram to green hydrogen's cost. That's why Kibria says fiddling with electrolyzers doesn't make as much sense as simply attaching green hydrogen projects to desalination plants. "We don't need to reinvent the wheel," he says. "This is a solved problem."

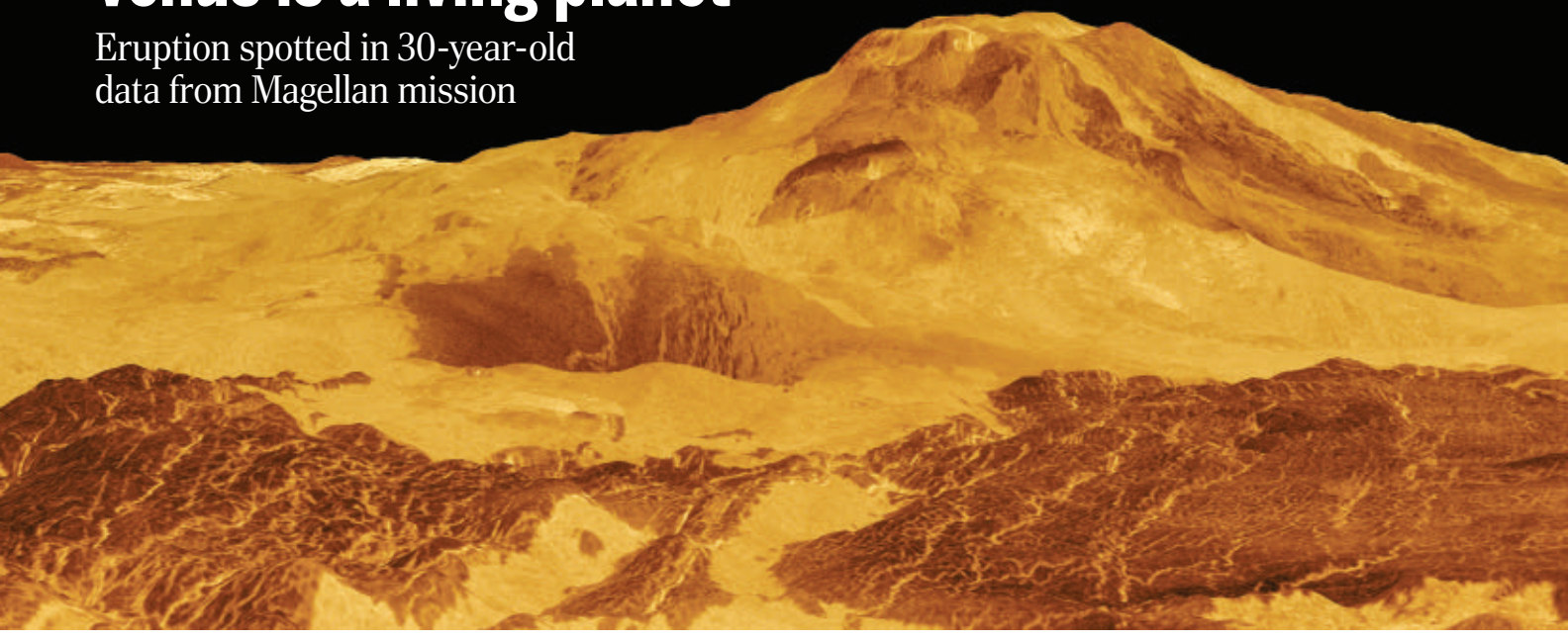
Mahmood disagrees. For starters, he says, desalination isn't a ready option for countries that can't afford large-scale capital projects. Moreover, he says, corrosion-resistant electrodes may also be useful for tapping other impure water sources, such as wastewater and brackish water. "We need to keep working on alternative technologies," he says. ■

PHOTO: ANGEL GARCIA/BLOOMBERG VIA GETTY IMAGES

PLANETARY SCIENCE

Active volcano shows Venus is a living planet

Eruption spotted in 30-year-old data from Magellan mission



By **Paul Voosen**

Choked by a smog of sulfuric acid and scorched by temperatures hot enough to melt lead, the surface of Venus is sure to be lifeless. For decades, researchers also thought the planet itself was dead, capped by a thick, stagnant lid of crust and unaltered by active rifts or volcanoes. But hints of volcanism have mounted recently, and now comes the best one yet: direct evidence for an eruption. Geologically, at least, Venus is alive.

The discovery comes from NASA's Magellan spacecraft, which orbited Venus some 30 years ago and used radar to peer through the thick clouds. Images made 8 months apart show a volcano's circular mouth, or caldera, growing dramatically in a sudden collapse. On Earth, such collapses occur when magma that had supported the caldera vents or drains away, as happened during a 2018 eruption at Hawaii's Kilauea volcano. "I'm totally tickled, as a geomorphologist, to see this," says Martha Gilmore, a planetary scientist at Wesleyan University who was not involved in the study.

Witnessing this unrest during the short observation period suggests either Magellan was spectacularly lucky, or, like Earth, Venus has many volcanoes spouting off regularly, says Robert Herrick, a planetary scientist at the University of Alaska, Fairbanks. Herrick, who led the study, says, "We can rule out that it's a dying planet."

The discovery, published this week in *Science* and presented at the Lunar and Planetary Science Conference, makes Venus only the third planetary body in the Solar System with active magma volcanoes, joining Earth and Io, Jupiter's fiery moon. It means future missions to Venus will be able to study "bare, gorgeous new rock" that provides a sample of the planet's interior, Gilmore says. The discovery of more volcanoes, in old or future data, will also help scientists understand how Venus is shedding its interior heat and evolving. And it will shake scientists out of their long-standing view that a spasm of activity a half-billion years ago repaved the planet's surface—as evidenced by a relative paucity of impact craters—and was followed by a long period of quiet. "We've all had our stagnant-lid lenses on to understand the planet," says Suzanne Smrekar, a planetary scientist at NASA's Jet Propulsion Laboratory (JPL). "We're finally getting an eye correction."

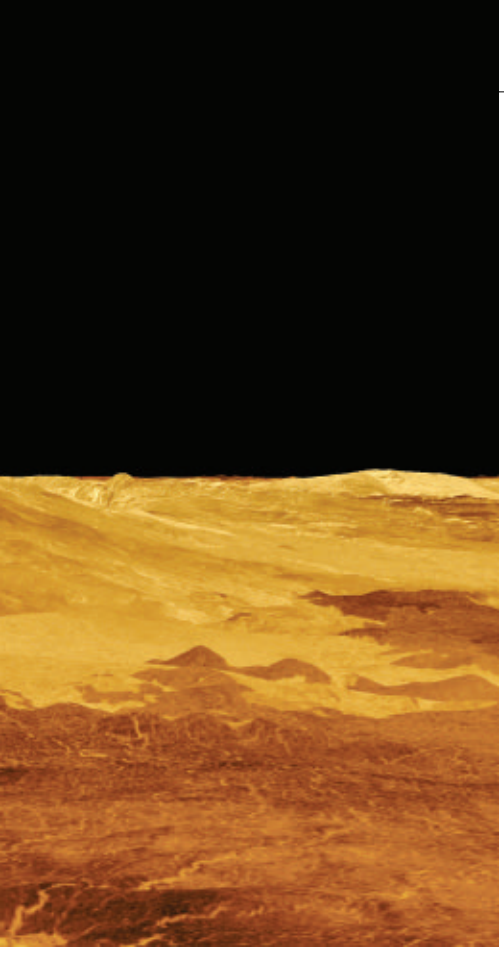
Recent years had brought hints that Venus has some geologic life. In 2010, researchers on the European Space Agency's Venus Express mission detected three anomalously hot regions, which they interpreted as lava flows a few million years old that hadn't yet cooled off. A couple of years later, the spacecraft found atmospheric spikes of sulfur dioxide, suggesting it was supplied by a variable source, such as volcanoes. And in 2021, a reanalysis of Magellan data indicated large blocks of crust had been jostled around like pack ice—a sign of rock stirring below the surface.

Prompted by these hints, Herrick decided to take another look at the Magellan data. "It's essentially looking for a needle in a haystack with no guarantee there's a needle there," he says. He targeted obvious candidates, such as Maat Mons, a volcano taller than Mount Everest. Magellan had already found that the force of gravity above it was surprisingly low—a sign that a hot plume of less-dense rock from the mantle might be fueling it, like the plume that sits beneath Hawaii. And microwave radiation from the summit suggested its surface had the chemistry of fresh lava.

Herrick had an unlikely ally in his search: endless pandemic Zoom meetings, which gave him time to compare radar images made at different times. "If anyone asks about a specific meeting, I was fully attentive at that one," he jokes.

The hunt was hard. At a resolution of several hundred meters, Magellan images are relatively coarse, only sensitive to the biggest changes in the landscape. Moreover, during its 5-year mission, the spacecraft revisited the same spots at most three times, and during its second campaign, its radar had been ro-

IMAGE: NASA/JPL



Maat Mons, 9 kilometers high, is Venus's tallest volcano. A collapse of its caldera signaled an eruption.

tated 180°. Comparing ground features from opposite angles is far from intuitive, Herrick says. "The same things look quite different."

But after hundreds of hours of tedious comparisons, covering less than 2% of the venusian surface, Herrick spotted what looked like a changed caldera. To avoid being fooled, he contacted Scott Hensley, a radar specialist at JPL well-known for debunking past false alarms in Magellan data. Hensley modeled what an unchanged caldera should have looked like during the second Magellan pass—starkly different from what was observed. The second image also appeared to show fresh lava flows, but those could have been hidden from view during the first pass, Herrick cautions. Still, the caldera changes are unequivocal evidence of volcanic activity, Smrekar says.

The discovery is just a preview of what is likely to come with three new Venus missions due to launch in the next decade: the European EnVision orbiter and NASA's DAVINCI and VERITAS missions. Both EnVision and VERITAS will be equipped with sharper radar vision than Magellan, making them well suited to monitoring the burps and twitches of a living planet, Herrick says. "We're guaranteed to see some really big changes." ■

INFECTIOUS DISEASE

White House budget includes push to eliminate hepatitis C

$11.3 billion would go to treatment, testing, and education

By **Mitch Leslie**

It's a minuscule part of the Biden administration's fiscal year 2024 budget proposal, announced last week (see story, p. 1078). But a 5-year, $11.3 billion program has a big ambition: to eliminate the deadly liver disease hepatitis C from the United States.

"I can't really recall a circumstance quite like this, where we have the chance to do something this groundbreaking," Francis Collins, onetime science adviser to President Joe Biden and former head of the National Institutes of Health (NIH), said in an interview with *JAMA*, which also published an editorial Collins co-authored advocating the proposal. If funded by Congress, the program would expand testing, broaden access to powerful antiviral drugs, and boost awareness.

"The field has been waiting for this for a long time," says transplant hepatologist David Kaplan of the University of Pennsylvania Perelman School of Medicine. Eliminating the disease "is possible and feasible," he says, noting that other countries are on their way to meeting that goal. Still, "It will be a challenge," says pediatric hepatologist James Squires of the UPMC Children's Hospital of Pittsburgh. "There's never been an eradication of an infectious virus without a vaccine."

> **"We can do it. Shame on us if we don't."**
> **William Balistreri**, University of Cincinnati

Hepatitis C kills more than 15,000 people in the United States every year. The virus that causes it spreads mainly through intravenous drug use and attacks the liver, often eventually causing cirrhosis, liver failure, and cancer. The Centers for Disease Control and Prevention estimates 2.4 million people in the United States harbor the virus. But the statistics are shaky, and pediatric hepatologist William Balistreri of the University of Cincinnati says the number could be as high as 10 million.

Despite the lack of a vaccine, researchers can talk seriously about elimination because drugs known as direct-acting antivirals (DAAs), first approved in the United States in 2013, are so effective. Just an 8- to 12-week course can oust the virus from more than 95% of patients. The introduction of DAAs spurred the World Health Organization (WHO) to make the elimination of hepatitis C by 2030 a goal. The disease wouldn't disappear, though. Instead, WHO aims to cut new cases by 90% and deaths by 65%.

A survey published earlier this year found that 11 countries were on track to meet those targets. One is Egypt, which slashed its disturbingly high infection rate by testing more than 50 million residents and treating 4 million. Australia, Japan, Georgia, and several nations in Europe have made similar progress. The United States has lagged because it lacks a national effort and, Kaplan says, getting treatment takes "too many steps."

For example, estimates suggest about 10% of the roughly 2 million people in the United States who are in jail or prison carry the virus, but they often go without testing and treatment. Other barriers also loom. In other countries, patients can undergo so-called point-of-care RNA tests at locations such as community health centers and substance abuse treatment clinics. If they test positive, they can receive treatment on the same visit. But in the United States, the tests have to be processed at off-site labs, forcing patients to return for their results and delaying treatment.

The new program would accelerate approval of point-of-care RNA tests. It would also tackle one of the biggest treatment obstacles—drug costs. Although the price of DAAs has fallen by about 75% since they were introduced, a full course still runs about $20,000. To improve treatment for underserved populations, the program would adopt the so-called subscription, or Netflix, model, first tested by Louisiana, in which the government pays drug companies a set amount for as much drug as it needs, rather than paying per dose.

All that may not be enough to eliminate hepatitis C, Kaplan says, but "it will make a significant dent in the problem." Still, the effort needs approval from Congress, including a Republican-controlled House of Representatives intent on slashing federal spending. Failing to seize this opportunity would be a huge loss, Balistreri says. "We can do it. Shame on us if we don't." ■

U.S. SCIENCE POLICY

Biden backs science in 2024 spending blueprint

But Congress will have its own views on setting funding levels for research agencies

By **Jeffrey Mervis**

Conquering cancer, commercializing research, and achieving greater equity in funding rank as top science priorities in a $6.8 trillion budget that U.S. President Joe Biden submitted to Congress last week. The proposal would expand the National Science Foundation (NSF) but would hold spending at most units of the National Institutes of Health (NIH) to current levels.

As with all budget requests, Biden's blueprint for 2024 is simply the starting point for negotiations with legislators over everything from taxes to countering China's growing economic and military might. And with both parties pushing for bigger defense budgets and Republicans vowing to reduce federal spending, the prospects for expanding domestic discretionary programs, which include all nondefense research activities, remain uncertain at best.

Arati Prabhakar, the president's science adviser and director of the White House Office of Science and Technology Policy, made a pitch for Biden's priorities at a 13 March White House event. "We need robust health and plentiful opportunity for everyone, we need to overcome the climate crisis, we need a competitive economy that creates jobs, and we need to maintain global security and stability," Prabhakar said. "And the purpose of American R&D is to make [those things] possible."

Biden's request would boost current R&D spending by 4%, to a record $210 billion. But basic research would rise by only 2%, to $89 billion, and late-stage development of military systems would receive more than half of the overall $9 billion increase.

"Every budget happens in a constrained environment," Prabhakar acknowledged. And that means the proposal has winners and losers (see table, above).

Biden's Cancer Moonshot, which aims to reduce cancer deaths by 50% over 25 years, gets top billing in the budget's research chapter, along with the new Advanced Research Projects Agency for Health, meant to accelerate cures for diseases. In contrast, spending at most of NIH's 27 institutes and centers would remain flat if Congress matches the overall 1.9% boost sought for NIH.

"FASEB [the Federation of American Societies for Experimental Biology] is very disappointed that the ... request for NIH does not even meet biomedical research inflation, let alone invest in promising new areas of science," says Jennifer Zeitzer of FASEB.

Many agencies had not released detailed budget plans when *Science* went to press, but the initial numbers included sizable top-line boosts for some. Laura Kolton, who leads the Science Coalition, said the group was pleased that the request "prioritizes robust, sustained investment for key research programs."

The 2024 request comes on the heels of two major laws enacted last year that increased spending at several science agencies, notably NSF, the National Institute of Standards and Technology, and the Department of Energy's (DOE's) Office of Science. One of those bills also endorsed—but did not appropriate—even bigger increases for this year and beyond at NSF and DOE science.

The Biden budget falls far short of meeting those expectations, however. For example, Biden's requested $11.3 billion for NSF falls $4.2 billion short of the spending Congress authorized in 2024 under the CHIPS and Science Act.

"With global competition continuing to grow, particularly in the scientific arena, it's critical that Congress back proposed research investments in the CHIPS and Science Act with real dollars that drive American innovation," Mark Becker, president of the Association of Public and Land-grant Universities, said in a statement.

Biden also proposes just one-third the money Congress authorized for NSF's new Technology, Innovation, and Partnerships (TIP) directorate. TIP was created last year to help researchers turn their discoveries into marketable products and new industries, and its flagship program to support regional innovation centers would grow by 50%, to $300 million, if Biden has his way.

The centers are designed to spread NSF's research dollars more evenly across the country, a goal that Congress mandated in the CHIPS act. NSF wants to serve the "missing millions" now excluded from NSF's research and training programs, says its director, Sethuraman Panchanathan. Biden also requested a 15% boost, to $281 million, in NSF's long-running program to boost the ability of have-not states to compete for NSF funds.

DOE's Office of Science, its basic research wing, would grow by $680 million, to roughly $8.8 billion. The 8% increase marks a step toward realizing the nearly 50% increase over 5 years authorized by the CHIPS act. And national and global programs aimed at curbing climate change would get $16.5 billion spread across numerous agencies, according to the request, including $5.1 billion for research.

But fierce opposition from Republicans, who now control the House of Representatives, could doom those requested increases. "This budget proposal boasts about spending taxpayer dollars on international climate slush funds ... while shortchanging the basic research that has been proven to advance our economy, lower energy prices, and reduce greenhouse gas emissions," said Representative Frank Lucas (R–OK), chair of the House science committee.

Spending hawks also noted that, although Biden's spending blueprint forecasts a $3 trillion drop in the federal deficit over the next decade, it would raise that deficit by $1 trillion next year if enacted. ■

Going up?

Some research programs would grow significantly in 2024 if President Joe Biden gets his way.

	FY24 REQUEST	INCREASE OVER FY23
National Cancer Institute	$7.8 billion	$500 million
NIH precision psychiatric initiative	$200 million	New
Advanced Research Projects Agency for Health	$2.5 billion	$1 billion
NSF technology directorate (TIP)	$1.2 billion	$305 million
TIP regional innovation engines	$300 million	$100 million
Graduate research fellowships	$380 million	$58 million
NASA earth science	$2.15 billion	$305 million
Fusion research	$1 billion	$332 million
Advanced Research Projects Agency for Energy	$680 million	$210 million
National Institute of Food and Agriculture grants	$550 million	$95 million
National Center for Education Research	$75 million	New

People take refuge in a cooling center in Portland, Oregon, during a June 2021 heat wave that struck the Pacific Northwest, killing hundreds.

PHOTO: KATHRYN ELSESSER/AFP VIA GETTY IMAGES

ENVIRONMENTAL HEALTH

Schizophrenia pinpointed as a key factor in heat deaths

The mental illness tripled the risk of death during a searing 2021 heat wave, researchers find

By **Warren Cornwall**

On 25 June 2021, as a blanket of hot air descended on the Pacific Northwest, British Columbia's provincial government issued a news release warning about the approaching heat wave's dangers. The announcement drew attention to the elderly, children, people working or exercising outdoors, pets, and "people with emotional or mental health issues whose judgement may be impaired."

Even so, more than 600 people died from the heat in British Columbia, as temperatures topped 40°C for days, shattering records in a region better known for temperatures usually half as high.

Now, new research has zeroed in on one of the hardest hit groups: people with schizophrenia. Epidemiologists combing through provincial health records found that, overall, those with mental health conditions seemed to have an elevated risk of a heat-related death. That was most severe for people with schizophrenia—a 200% increase compared with typical summers. "Those are really large numbers and … alarming," says Peter Crank, a geographer at Oklahoma State University, Stillwater.

"We didn't protect them," laments Sarah Henderson, an environmental epidemiologist at the British Columbia Centre for Disease Control who oversaw the research, published on 15 March in the journal *GeoHealth*. "These results show that people with schizophrenia need extra protection, extra support, and extra care."

Earlier research had shown schizophrenia can make people more vulnerable to heat. Crank, for instance, recently reported a link between higher temperatures and hospitalization of people with schizophrenia in Phoenix. But the connection "just hasn't made it to the mainstream," Henderson says.

To examine which chronic health problems put people at greater risk during the heat wave, Henderson and her team paired a detailed accounting of deaths in British Columbia with medical records in Canada's national health care system. They compared the medical histories of 1614 people who died during the 8-day heat wave with 6524 deaths on the same dates during the prior 9 years. The data covered 26 medical conditions, from heart disease to dementia to osteoporosis.

Henderson expected the data to confirm the widespread belief that kidney disease and heart disease are key risk factors during extreme heat. The records did show people with coronary artery disease were 18% more likely to die during the 2021 heat wave than in previous years, and those with kidney disease were 36% more likely to die. But the increase in schizophrenia deaths dwarfed both those conditions.

Overall, more than 8% of the people who died during the hot week had a history of schizophrenia, compared with 2.7% in the same week during a typical year. The results were even more striking for a subset of the total deaths—the 280 that the provincial coroner's service certified as being heat related. Thirty-seven people who died—more than 13%—had schizophrenia.

The death toll isn't a surprise to George Keepers, a psychiatrist and schizophrenia specialist at Oregon Health and Science University who wasn't involved in the study. "There's a whole host of things that people with this very unfortunate illness are vulnerable to," Keepers says.

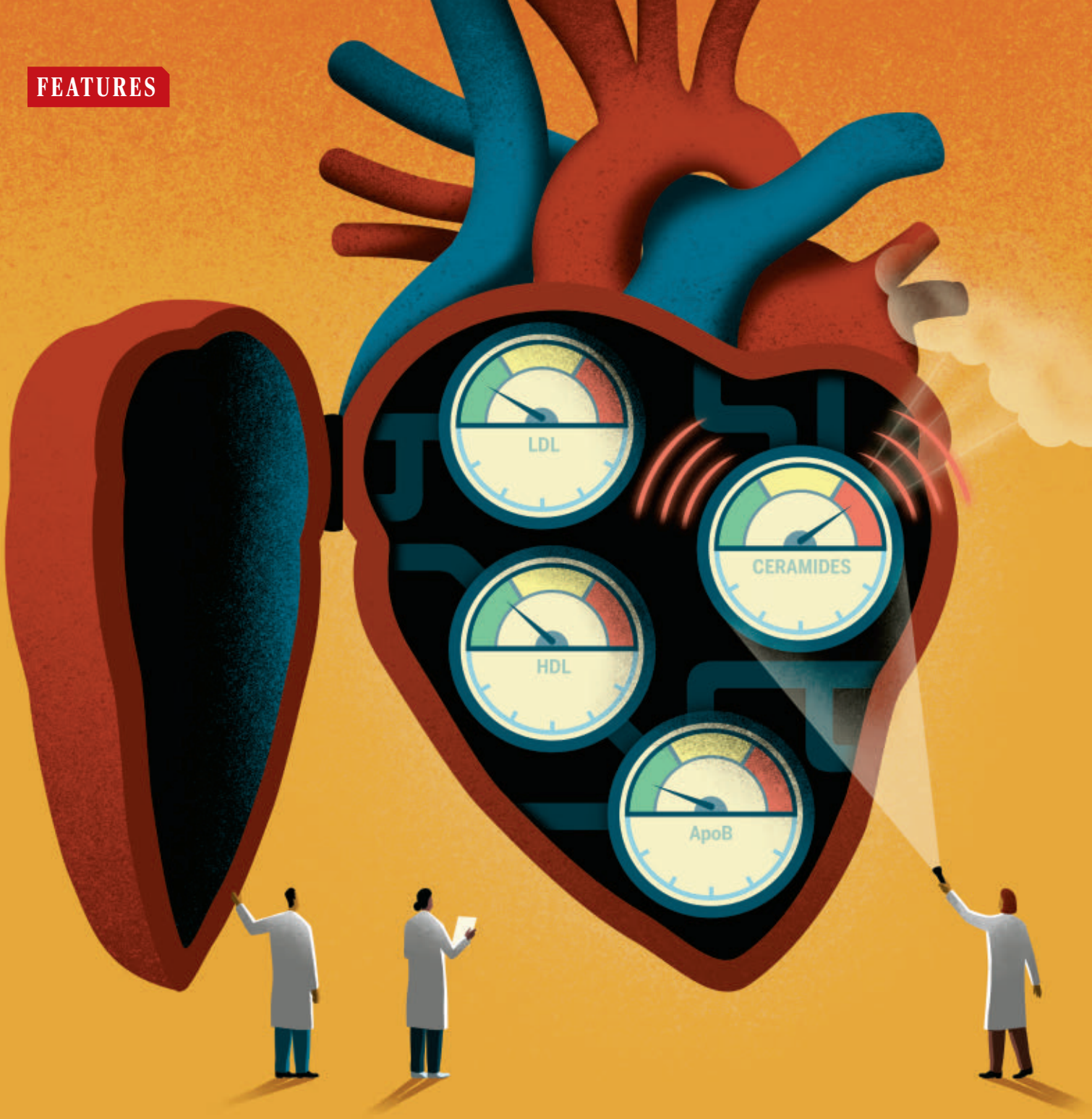
For instance, schizophrenia can affect the brain's hypothalamus, which helps regulate temperature through sweating and shivering. Some antipsychotic medications can raise body temperature, which can have deadly effects when coupled with extreme heat. The disease affects people's ability to make reasoned decisions or sense when they are ill. People with schizophrenia tend to have other conditions tied to heat-related illness, such as diabetes. Finally, schizophrenia is associated with isolation and homelessness, which puts people at risk when temperatures rise.

In Phoenix, the first U.S. city to create an office for addressing heat risks, "the word schizophrenia does not appear in the … heat response plan. And maybe it should," says David Hondula, who leads the city's Office of Heat Response and Mitigation. He notes that local data show people without housing—some of whom have schizophrenia—are two to three times more likely to die from heat than the overall population.

The findings in Canada have already prompted more research. Liv Yoon, a sociologist at the University of British Columbia, Vancouver, is preparing to delve into the stories of people with schizophrenia who survived the heat wave. "We realize there's more going on than simply the physiological mechanism," Yoon says. She hopes talking with survivors will shed light on social factors contributing to the surge in deaths.

As scientists warn that climate change will bring more deadly heat waves, the nonprofit British Columbia Schizophrenia Society has ramped up efforts to educate caregivers about the danger, says CEO Faydra Aldridge. "I don't think any of us were as prepared in any area for this heat wave that happened," she says. "Now, we're much more aware of the potential risks for people living with schizophrenia."

Henderson, meanwhile, chairs a provincial committee formed after the 2021 heat wave to prepare for future events. Public announcements should make much more explicit warnings, she says. "When we're talking about risk factors for extreme hot weather, schizophrenia needs to be near the top of the list." ■



STRAIGHT FROM THE HEART

Lipids called ceramides may be better predictors of cardiovascular problems than cholesterol. Doctors and pharma are waking up to their potential

By **Mitch Leslie**

Stephanie Blendermann, 65, had good reason to worry about heart disease. Three of her sisters died in their 40s or early 50s from heart attacks, and her father needed surgery to bypass clogged arteries. She also suffered from an autoimmune disorder that results in chronic inflammation and boosts the odds of developing cardiovascular illnesses. "I have an interesting medical chart," says Blendermann, a real estate agent in Prior Lake, Minnesota.

Yet Blendermann's routine lab results weren't alarming. At checkups, her low-density lipoprotein (LDL), or "bad," cholesterol hovered around the 100 milligrams-per-deciliter cutoff for normal values, and her total cholesterol—the good and bad versions combined—remained in the recommended range. "I thought I was cruising along just fine," she says.

But because Blendermann's risk was unclear, in late 2021 her doctor decided to refer her to cardiologist Vlad Vasile at the Mayo Clinic. To pin down her susceptibility to atherosclerosis, Vasile prescribed a test for substances Blendermann had never heard of: lipids called ceramides. Long overlooked, they are emerging as powerful alternatives to standard markers of heart disease risk such as LDL cholesterol. Blendermann's score was moderately high, suggesting that compared with a person with a low score, she was more than twice as likely to suffer a cardiovascular event such as a heart attack. "It woke us up big time," she says. "The ceramides told me the bigger story." She began to take cholesterol-lowering drugs and overhauled her diet and exercise regime.

Doctors and drug companies are also warming to the medical possibilities of ceramides. Blendermann is one of just a few thousand people in the United States to have undergone ceramide blood testing, which is only performed by the Mayo Clinic. But later this year, lab testing giant Quest Diagnostics will start to offer the analysis, potentially making it available to many more patients.

The first drugs specifically designed to lower ceramide levels are also on the horizon, with at least two companies hoping to begin clinical trials within the next year or so. And researchers are refining their picture of how these molecules, which account for less than 1% of the lipids in the body, exert such a powerful influence over our physiology. Ceramides are essential for a variety of cellular functions. But a stack of studies also implicates high levels of the molecules in heart disease and illnesses such as diabetes and fatty liver disease, suggesting they may cause havoc as well.

"There is overwhelming evidence that [ceramides] are major driving forces for metabolic dysfunction," says physiologist Philipp Scherer of the University of Texas Southwestern Medical Center. That makes them valuable for assessing patients' odds of developing some chronic illnesses—and "an excellent predictor of cardiovascular risk," says Jeff Meeusen, co-director of cardiovascular laboratory medicine at the Mayo Clinic.

> **Ceramides "are major driving forces for metabolic dysfunction."**
>
> **Philipp Scherer**, University of Texas Southwestern Medical Center

Still, the medical community has not embraced ceramides. Before that happens, cardiologists will have to accept an unfamiliar test and learn how to interpret the results alongside standard risk factors. And before patients start to receive ceramide-lowering drugs, developers will have to show that interfering with compounds fundamental to the body does more good than harm.

UNTIL A LITTLE OVER 30 years ago, ceramides "were not on anyone's radar screen," says Yusuf Hannun, a lipidologist at Stony Brook University. The few researchers who did think about the molecules, which are found throughout the body, assumed they were metabolically inert. In 1993, Hannun and his colleagues performed one of the first studies that helped change that perception.

The researchers wanted to find out how a specific immune system molecule spurs malignant cells to commit suicide, protecting against cancer. They discovered the molecule acts through ceramides, suggesting the lipids are important for conveying messages within cells. Soon afterward, a new technique called liquid chromatography-mass spectrometry revolutionized the study of the lipids. The technique, which can sort complex molecular mixtures, revealed that cells carry numerous ceramide varieties—mammals boast more than 200 types—and scientists have been trying to tease out the molecules' functions ever since.

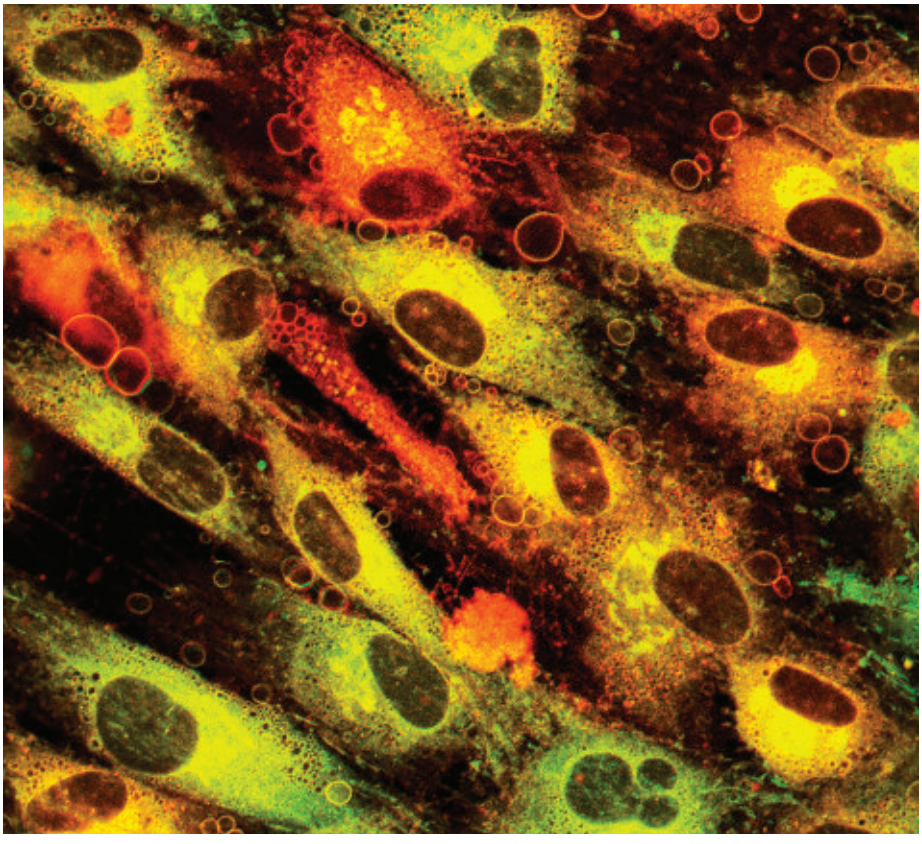
One place the lipids are essential, says biochemist Ashley Cowart of Virginia Commonwealth University, is the skin, which "has a very diverse ceramide population." There, they help maintain a solid protective layer—that's why skin cream–makers load their products with synthetic ceramides or those derived from natural sources. In the skin and elsewhere in the body, cells incorporate different types of ceramides to fine-tune the fluidity of their outer membranes, which influences cellular functions such as movement, division, and communication. Ceramides also serve as raw materials for the synthesis of other lipids. In short, says lipid biochemist Tony Futerman of the Weizmann Institute of Science, "We can't survive without ceramides."

But as researchers have discovered, ceramides can also turn against us. They can infiltrate the lining of blood vessels and usher in LDL cholesterol particles, thus contributing to atherosclerosis. They can inhibit production of nitric oxide, a chemical messenger that relaxes artery walls and helps keep the vessels open. Some ceramides appear to promote insulin resistance, a defect in sugar metabolism characteristic of type 2 diabetes and other conditions. The molecules can also reduce energy production by mitochondria, the organelles that provide cells' chemical fuel. And the cell suicide that ceramides can trigger, although protective against cancer, may damage healthy tissue in organs such as the heart.

Why do ceramides sometimes go bad? Some are born that way. A particular ceramide's character depends on the size of its acyl tail, a portion of the molecule that can contain from 12 to more than 26 carbons. "The length of the acyl chain has enormous importance in cell physiology and in cell pathophysiology," Futerman says. In general, ceramide varieties with long tails are more damaging, and certain molecules with 16-, 18-, or 24-carbon tails may be the most dangerous, for reasons yet unknown.

Ceramides may also become deleterious when our bodies produce too much of them. We break down the fats we eat to yield fatty

ILLUSTRATION: (OPPOSITE PAGE) DAVIDE BONAZZI/SALZMAN ART



The bubbles in this image of skin cells are rich in ceramides. The lipids help maintain the integrity of the skin's outer layer—which is why ceramides are included in skin creams.

PHOTO: VSHYUKOVA/SCIENCE SOURCE

acids, some of which get shuttled into the pathway that produces ceramides. Our cells normally only manufacture small amounts of ceramides. When our diet contains too much fat, however, synthesis of the molecules booms. "The ceramide pathway is kind of a spillover pathway" for excess fatty acids, Scherer says.

The link to diet likely explains why ceramides surge in so many diet-related metabolic conditions. For instance, researchers using liquid chromatography-mass spectrometry have found elevated levels of specific ceramides in patients with obesity, type 2 diabetes, nonalcoholic fatty liver disease, and several types of cardiovascular conditions, including atherosclerosis, heart failure, and stroke. And rodent studies suggest ceramides may be more than just bystanders. Using chemical treatments or genetic manipulations to cut ceramide levels can protect the animals from many of these ailments.

Some researchers remain unconvinced. "Whether they are causative or a result—in my view, we don't know," Futerman says. But physiologist Scott Summers of the University of Utah, who has been studying ceramides for more than 20 years, is one of the researchers who accepts their health effects. "The data for us have been perfectly clear that these are important molecules."

RESEARCHERS CONTINUE to dig deeper into the biology of ceramides, but they are also eyeing the lipids as potentially valuable biomarkers to gauge a patient's heart disease risk. The traditional factors for assessing this risk include age, sex, whether the patient smokes or has diabetes, and lab measurements of lipids such as LDL cholesterol. However, these indicators don't flag everyone who is in danger. In fact, about 15% of people who suffer heart attacks have no standard risk factors at all.

Ceramides may fill the gap. In one 2016 study, clinical pharmacologist Reijo Laaksonen of Zora Biosciences and Tampere University and colleagues analyzed cholesterol and ceramide levels in people with heart disease. Blood ceramides accurately forecast whether these people would die from heart attacks. For example, the abundance of one ceramide variety with a 16-carbon tail was 17% higher in patients who perished than in individuals who survived. In contrast, LDL cholesterol provided no insight—it was higher in the people who didn't have heart attacks, the scientists reported. Laaksonen and his colleagues, as well as other research teams, have also found that ceramide levels reveal cardiovascular risk in the general population. Overall, studies on more than 100,000 people confirm the predictive power of ceramide testing, Laaksonen says. "It's very fair to say the ceramide test is the best lipid-based risk marker for cardiovascular events." Zora has licensed its ceramide scoring algorithms to the Mayo Clinic and Quest.

Meeusen says he and his Mayo Clinic colleagues are generally wary of new medical tests, but that the evidence for ceramide testing was compelling enough to start offering the assays to patients in 2016. The team was also swayed by research suggesting ceramides are involved in cardiovascular disease development. "Ceramides [are] more directly involved with atherosclerosis progression compared to cholesterol," Meeusen says.

DESPITE THOSE ADVANTAGES, ceramide testing remains limited. Meeusen says the Mayo Clinic performs about 1000 of the analyses per month, mostly in-house requests. In comparison, the clinic performs several times that many standard lipid panels every day.

Other providers are beginning to offer ceramide testing as well, however. For example, most private clinics and about one-half of public hospitals in Finland do so, Laaksonen says. Quest's imminent entry into the market will further increase availability.

Marc Penn, medical director for Quest's Cardio Metabolic Endocrine Franchise, says the company decided to offer ceramide tests because they are essentially three tests in one. For most patients today, Penn says, doctors assemble a fragmentary picture of their risk for conditions such as heart disease and diabetes by performing separate tests for lipids, blood sugar, and inflammation. But measuring ceramides provides a comprehensive assessment of a patient's risk for metabolic diseases because all three factors affect the levels, he says.

Nobody expects ceramide testing to usurp the standard lipid panel. A ceramide test is more complex to perform because it requires mass spectrometry, which is not available in most clinical labs. It is also

Do-it-all molecules

Ceramides can raise the risk of disease—but when they are present at normal levels, they play critical roles in the body.

- Seal outer layer of skin
- Trigger suicide of cells
- Control cell membrane fluidity
- Stimulate internal cellular recycling
- Provide substrates for synthesis of complex lipids

about 10 times more expensive, running around $100 at the Mayo Clinic. Moreover, it remains to be seen how many practicing cardiologists will opt for the tests even once they're easier to order.

Neha Pagidipati, a preventive cardiologist at Duke Health, says she is open to the idea. "There is a place for additional measurements to understand who is at risk for cardiovascular disease." Still, she says that although one of her patients asked about ceramide testing, she has never ordered it and remains unsure about its clinical value. "It needs to be clearer what I'd advise my patients to do with that information."

Summers worries some recommendations based on ceramide results could be counterproductive. Researchers have noted that blood ceramide levels tend to fall after patients improve their diet, exercise more, or take cholesterol-lowering medications such as statins. Recommending exercise is probably safe, Summers says, but statins "might just be keeping [ceramides] in the liver, where they do a lot of their damage." What's missing are data from clinical trials in which researchers test whether interventions such as diet and lipid-lowering treatments not only reduce ceramide levels, but also translate into improved health.

In 2020, Laaksonen and colleagues launched the first trial that will try to address that omission. The researchers are identifying 2000 patients with heart disease who have high levels of ceramides and three other biomarkers of cardiovascular risk. One-half of the patients will enter an intensive program, receiving twice-yearly coaching sessions about diet and exercise and frequent advice from a smartphone app. They will also get tailored recommendations for blood sugar– and lipid-lowering drugs. The other half of the group will receive regular care from their physicians. The researchers plan to follow the participants for 3 years, measuring their rates of cardiovascular events, to determine whether the more aggressive approach provides disease protection in addition to reducing ceramide levels.

ALTHOUGH DIET AND EXERCISE may reduce ceramide levels, some researchers have sought a more direct approach: drugs that disrupt ceramide synthesis or break down the molecules. So far, big pharmaceutical companies' efforts to develop such drugs have faltered for various reasons. In the early 2010s, for instance, researchers at Eli Lilly and Company identified two compounds that block the enzyme SPT, which catalyzes the first step in ceramide synthesis. These molecules slashed ceramide levels in rodents by 60% to 80%. But they also caused the lining of the animals' intestines to peel off, leading the company to kill further development.

Biotechs are now picking up where big pharma left off, Scherer says. The company that Summers co-founded in 2016, Centaurus Therapeutics, has crafted a molecule that inhibits DES1, the enzyme that catalyzes the final step in ceramide synthesis. Summers says blocking this enzyme is likely to be safer than targeting SPT, noting that his team deleted the gene for DES1 in rodents without serious side effects. Centaurus is now amassing the animal safety data the U.S. Food and Drug Administration (FDA) requires to greenlight a clinical trial, says Jeremy Blitzer, the company's chief scientific officer. He wouldn't speculate on a start date, but says, "We are on a short path to a first dose in humans."

Another biotech, Aceragen, is probing a different compound that breaks down ceramides and plans to begin a clinical trial within a year. The company intends to test the drug for patients with a rare and often-fatal metabolic condition called Farber disease, which results in abnormally high ceramide levels.

Other researchers are pursuing different strategies for reducing ceramide concentrations, but their work is at an earlier stage. Cardiologist Christian Schulze of the University of Jena and colleagues are trying to replicate the effects of a drug known as myriocin, which cuts ceramide levels dramatically in mice, protecting them from heart failure, slowing atherosclerosis, and improving insulin sensitivity. The catch is that myriocin, which was isolated from a fungus, suppresses the immune system, which once made it a potential treatment for rejection of organ transplants. "The side effects are what it was developed for," Schulze says. But immune suppression boosts vulnerability to infections.

Using the crystal structure of myriocin's active site as a template, Schulze and his colleagues have developed several molecules that seem to trigger the same benefits without undermining immunity. They have tested these compounds in cells and plan to move on to rodent studies. Laaksonen and his colleagues have reached about the same stage with their work. They are aiming to reduce ceramide levels with short interfering RNAs, which diminish levels of specific proteins necessary for ceramide synthesis.

Whether these efforts will deliver practical anticeramide drugs remains to be seen. But patients like Blendermann are already benefiting from ceramides' power as risk markers. After getting her test result, she began to exercise more and eat more green vegetables and leaner meats such as fish and chicken. "That was huge for me. I grew up in a meat and potatoes family," she says. After 1 year, her ceramide score had plunged from eight to one, the second-lowest risk level. Her other lipids, including LDL cholesterol and total cholesterol, also improved. She credits the ceramide test with making her realize "I've got to get busy and get this right." ■

Blood drawn from patients isn't routinely tested for ceramides, but that could change as research underscores the power of these lipids for revealing susceptibility to heart attacks and metabolic diseases.

PHOTO: DMPHOTO/ISTOCK.COM

INSIGHTS

POLICY FORUM

CLIMATE COOPERATION

Lessons from China's overseas coal exit and domestic support

This dichotomy can inform environmental cooperation

By **Christoph Nedopil**

China achieved an important climate milestone in September 2021 when it unilaterally announced that it would stop building new coal-fired power plants abroad. A belief that this was driven by climate considerations and international pressure overlooks a distinctive dichotomy: China announced its overseas coal exit while not changing its basic approach to domestic coal plants. This is in the reverse order of other countries (e.g., Germany announced a domestic coal exit law, then a stop to public funding for overseas coal) and seems incongruent because both China's overseas and domestic coal investments depend on the same institutions and enterprises for finance and top-down policy signals and support. Analysis of this dichotomy suggest drivers that are at odds with a common view of China's approach to the environment, that of top-down steering with bottom-up implementation (*1*) or "authoritarian environmentalism" (*2*). This raises important issues in climate governance and offers insights on how to cooperate with China on green development.

China has become the world's greatest source of greenhouse gas emissions, and its biological diversity at home exhibits

Fanhai International School of Finance, Fudan University, Shanghai, P. R. China. Email: nedopil@fudan.edu.cn

PHOTO: OWNGARDEN/GETTY IMAGES

A coal-fired power plant alongside solar panels in Shanghai illustrates the concurrent domestic push by the Chinese government for both coal and renewable energy.

high vulnerability; internationally, China's influence through trade, finance, and investment—not least through its Belt and Road Initiative (BRI)—is among the biggest in the world, with severe biodiversity and climate implications (*3*). Given the scope and scale of China's influence, reduction of global environmental risks, such as climate change and biodiversity loss, cannot succeed without transforming China's domestic and international economy.

DRIVERS OF CHINA'S COAL DICHOTOMY

China was the world's largest public financier of overseas coal plants between 2006 and 2021 (*4*), completing about 54 GW of coal-fired power plants in 20 countries from Vietnam to Pakistan and from South Africa to Bosnia and Herzegovina (*5*). Domestically, China operates 49% of the world's coal capacity of 2046 GW and has another 251 GW of coal-fired power plants under development (*5*).

China's success in building overseas coal plants rested on a perceived low-risk model. China used its considerable domestic coal expertise in what were frequently noncompetitive government-to-government agreements with host countries or competitive bids based on an "all-China" offer: Chinese power plant developers—all state-owned enterprises—were responsible for building, operating, and/or owning the coal plants. Overseas projects were most often financed through Chinese policy banks (e.g., China Development Bank and Exim Bank of China) or state-owned commercial banks (e.g., Industrial and Commercial Bank of China). To cover credit and other risks, Chinese financiers required insurance by China's government overseas export credit agency Sinosure, which itself required sovereign guarantees from the recipient country's government (e.g., to guarantee payments under the power purchasing agreement).

However, particularly since 2019, economic, social, and political risks in China's overseas markets changed demand considerably (see also table S1): Multiple recipient countries reevaluated their electricity needs because of lower-than-expected economic growth, which caused, for example, Egypt's proposed 6.6-GW Hamrawein plant to be shelved in April 2020 and Bangladesh to request that China reduce financing for agreed coal plants in February 2021. Other countries (e.g., Vietnam, Pakistan) announced a focus on green energy and accelerated their net-zero targets. Other recipient countries were confronted with domestic social pressure against coal, such as the decoalinize movement that caused Kenya's 1.2-GW Lamu coal-fired power plant to stop construction in June 2019. Broader social concerns became evident when 263 environmental nongovernmental organizations from around the world addressed a letter to China's Ministry of Commerce in April 2020 (*6*) asking them to reevaluate China's engagement in overseas coal-fired power plants.

Simultaneously, financial risks for overseas coal plants affected Chinese supply-side decision-making: Increased sovereign debt risk in many recipient countries, exacerbated by COVID-19, reduced Sinosure's ability to provide insurance for new plants. This, possibly, led to the withdrawal of financing for a USD$3 billion planned plant in Zimbabwe in June 2021 (the Zimbabwean owner was still seeking financing in 2022). Meanwhile, with more global investors reducing financing for coal, the financing cost for coal-fired power plants increased on average by 38% in the period from 2017 to 2020 (with reference to 2007 to 2010), compared with a decrease in financing cost of 24% for offshore wind and 12% for onshore wind (*7*). This affected China's ability to finance operations of overseas plants in contrast to the financing of domestic plants, for which China's central bank has substantially expanded support since 2021 (*8*) to cover losses of coal plant operators.

In addition, carbon pricing schemes proliferated, which added cost risks, particularly to new overseas coal-fired power plants: In 2021, 64 carbon pricing initiatives were in operation compared with 58 initiatives in 2020 and 47 in 2018 (*9*). By contrast, China's own emission trading system, introduced in 2021, is based on emission intensity (i.e., limits on tonnes of carbon dioxide per megawatt-hour) rather than cap and trade (i.e., limits on total emissions). This allows "efficient" coal plants to improve profitability by selling carbon allowances rather than having to buy them.

Finally, a 40% increase in coal price volatility and a 30% average price increase during the period from 2018 to 2021 (compared with 2015 to 2018) affected Chinese operators of all new and existing plants, except in countries with capped or controlled coal prices, including China (*10*).

Consequently, of the 51 coal-fired power plants outside China and supported by China that were announced between the second half of 2014 and the end of 2020, only one plant became operational. By contrast, mothballing (a stop in construction) and cancellations of plants accelerated, with 25 plants shelved and 8 cancelled, which amounts to a total announced capacity of about 56 GW (see the figure). No new overseas plant was announced in 2020—a year before China's official announcement.

The challenges in China's overseas coal engagement allowed Chinese-led institutions that were tasked with green development, such as the Belt and Road Initiative International Green Development Coalition (BRIGC) under the Ministry of Ecology and Environment, to test the willingness of central government decision-makers to exit overseas coal under the banner of "building a green BRI." In December 2020, the Ministry and selected government experts, including those from the National Development and Reform Commission and the China Banking and Insurance

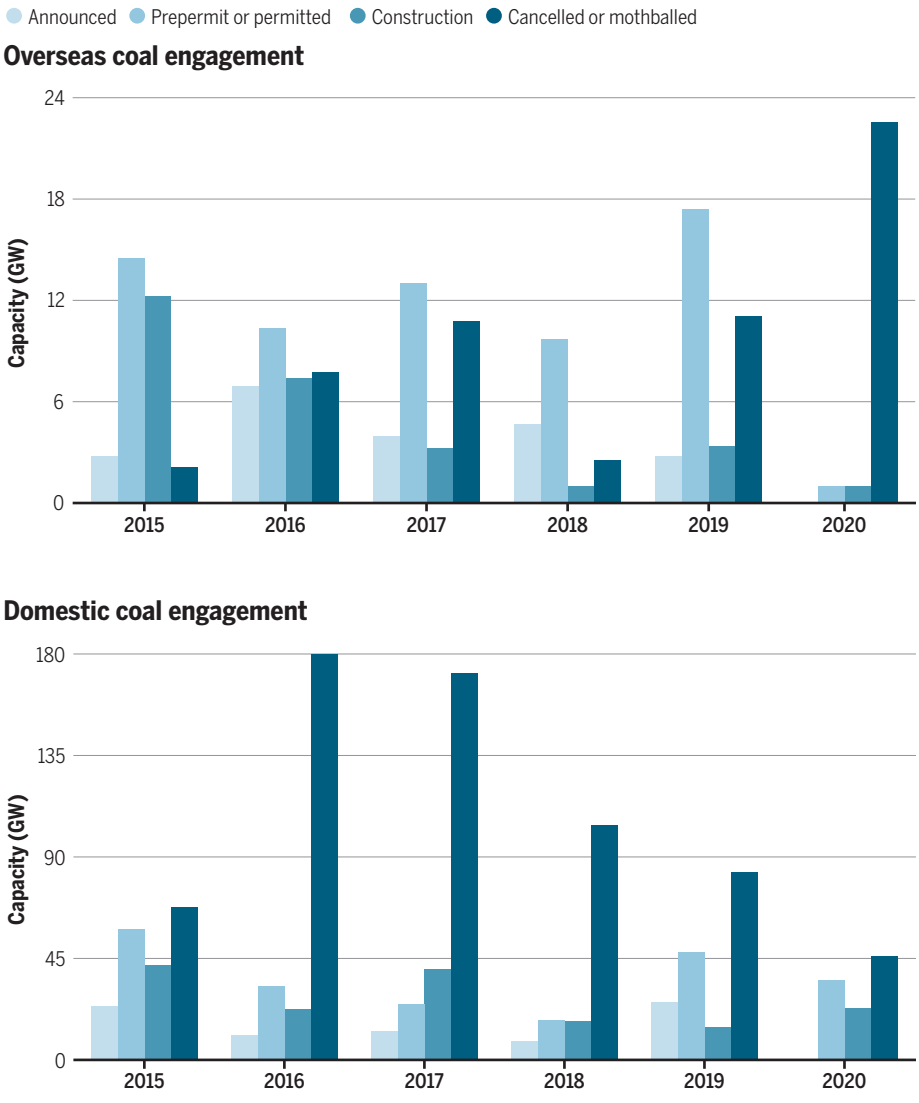
Regulatory Commission, backed the "Green Development Guidance for BRI Projects" that was developed with international support under the BRIGC, which labeled coal as "red" restricted projects (*11*). In February 2021, BRIGC completed a special research report that recommended a stop to overseas coal investments that was "submitted to the competent authorities and firmly underpinned the decision-making on China's overseas coal-related investments" (*12*).

Once China's political leadership made the decision to exit overseas coal, a goal seemed to be to maximize "green" soft power by providing a public top-level announcement instead of quietly reducing investments. The choice to announce at the United Nations General Assembly (UNGA) in September 2021 seemed superior for that soft-power goal compared with announcing at three other international leaders' forums in which China participated in 2021: at the annual Boao Forum hosted by China in April, which attracted mostly Chinese and Asian leaders and audience; at the G20 summit in October in Italy; and at the UN Climate Change Conference (COP26) in November in the United Kingdom, where China would have had to share reputational gains with the hosts. China's choice of the UNGA allowed China to present the decision as globally relevant and independent, compared with, for example, COP26, where it could have been seen as a "bargaining chip" or as being agreed to under external pressure.

In contrast to the foreign exit, China announced nonbinding ambitions to "gradually" reduce new coal plants in the future based on climate and energy policies ("1+N") to support China's 2060 carbon neutrality target. Practically, China expanded support of domestic coal as the "ballast stone" (baseload) to guarantee energy security after power outages in 2021, which was ironically induced by high coal prices and despite financial losses and stranded asset risks for most coal plant operators in China (*13*). Simultaneously, China rapidly expanded renewables (*3*).

Status change of Chinese-backed coal plants

Year-on-year status changes of Chinese-backed overseas coal plants and domestic coal plants are shown. Overseas coal-plants in 2020 with a capacity of 22.5 GW that had previously been announced, prepermitted, or under construction were cancelled or mothballed, whereas only 960 MW changed status from permitted to construction (top). Domestic plants that started construction increased from 14.3 GW in 2019 to 22.7 GW in 2020, whereas mothballed or cancelled plants decreased continuously from 180 GW in 2016 to 46 GW in 2020 (bottom).



GRAPHIC: K. FRANKLIN/*SCIENCE*

FINDINGS AND LESSONS

Although loopholes exist in China's overseas coal exit (*14*), the contrast with China's continued support for domestic coal suggests four findings. First, China's overseas coal exit seems to be based foremost on economic rather than long-term environmental considerations; otherwise, China would have also committed to stop building new coal plants at home. Second, political consensus for the exit was made possible by a relatively small coalition of proexit stakeholders supported by international partners. The domestic stakeholders had the ability to reach top leadership from within the system, whereas stakeholders in recipient countries had relatively little influence. This contrasts with a more complex stakeholder landscape that opposes a domestic coal exit involving political leaders, financial institutions, and state-owned enterprises along the whole supply chain (e.g., mining, transport, generation) across different Chinese provinces. This includes millions of workers in the domestic coal sector, compared with the limited number of Chinese workers affected by China's overseas coal exit. Third, this stakeholder complexity (and China's dependence on coal power) makes decision-making—and even policy recommendations—toward a domestic coal exit politically riskier, with the central government embracing a loss-aversion and low-conflict strategy that supports the expansion of both coal and renewables. Fourth, China seeks green soft-power gains and uses multilateral platforms for undivided attention for Chinese environmental progress (e.g., like the announcement for carbon neutrality in 2060 at the UNGA in 2020). Various identified risks and China's risk-control ability drive the dichotomy of China's domestic and international coal engagement (see the table).

Risk model for China's domestic and overseas coal engagement

Domestic risks may seem more controllable for Chinese authorities, compared with overseas risks

TYPE OF RISK	OVERSEAS COAL RISKS	PERCEIVED ABILITY TO CONTROL RISK	DOMESTIC COAL RISKS	PERCEIVED ABILITY TO CONTROL RISK
Economic	Demand (e.g., cancellation of plants) Financial (e.g., sovereign debt risks, cost of financing) Cost (e.g., cost of fuel, carbon price)	Low	Demand risks (controlled by central energy planning) Financial risks (controlled by central bank and banking regulator) Cost risks (controlled)	High
Social	Social resistance against coal (e.g., nongovernmental organization activism)	Low	Social activism limited	High
Political	Change of political parties or policies in partner countries	Low	Risk for domestic political stakeholders (e.g., to be responsible for social unrest, unemployment)	Medium
Reputational	Loss of reputation	Medium	Loss of reputation	Medium
Environmental	Environmental damages (climate, biodiversity)	Medium	Environmental damages (climate, biodiversity)	Medium

The variation in risks has implications for prioritizing areas of economic, policy, and technical cooperation on the environment. Importantly, these differ for domestic and overseas engagement. For domestic environmental cooperation, for example, China's domestic coal exit or biodiversity protection, engagement opportunities are more limited to political and reputational aspects, whereas, for example, international finance's impact would be limited in China's large economy. First, engagement should aim to reduce complexity and risk aversion of stakeholders by targeting stakeholders on the provincial or sector level (e.g., energy, finance) with more aligned needs and a higher risk appetite, rather than focusing on commitments from China's central government. Second, international partners can provide technical capacity and some financing in Chinese "pilot zones" with ecological mandates. This allows for targeted input of international ecological capacity and trust building between Chinese and international partners with the potential to scale experiences to other areas. Third, partners can expand knowledge diplomacy with Chinese institutions, which needs to incorporate the more-sensitive nature of domestic issues and focus on enabling local partners. By supporting exchanges between domestic and international institutions, for example, to research on just transition or green finance, international expertise can diffuse into Chinese domestic partners and thus policy deliberation.

For overseas engagement, for example, to support the retirement of China's overseas coal plants, economic, social, political, and reputational factors can be addressed. First, China's stakeholders would need to conclude that current nongreen engagement entails undue economic risk, for example, due to increased refinancing or operating costs (e.g., higher fuel prices, logistics prices) or even penalties (e.g., reduced financing from multilateral development banks). On the contrary, by providing support for the expansion of green activities or the reduction of harmful activities, for example, through buy-out financing like Asian Development Bank's Energy Transition Mechanism (*15*), the international community can increase overseas demand for green Chinese investments. Second, international partners can expand cooperation with civil society organizations in recipient countries to support local environmental advocacy. Third, in contrast to domestic knowledge diplomacy, international partners can expand collaboration with Chinese partners with central government influence or work in established multiparty collaboration forums [e.g., BRIGC, the China Council for International Cooperation on Environment and Development (CCICED)] to jointly develop green knowledge and policies. The Chinese partner can tailor the messages to Chinese political needs and policy frameworks (e.g., "ecological civilization," "green BRI") and deliver it internally to leadership.

Finally, for both domestic and international cooperation, the international community can constructively acknowledge specific areas of environmental progress, which would allow China to gain desired "green soft power," for example, through statements or study tours by international leaders (e.g., as seen in the international exploration of China's mobility electrification). This attention should also raise the bar against retrogression.

These findings on China's "Panda-Dragon" dichotomy in coal engagement highlight a more complex climate governance of China and should help develop more targeted cooperation strategies. ■

REFERENCES AND NOTES

1. A. Engels, *Palgrave Commun.* **4**, 101 (2018).
2. H. Yang *et al.*, *Nat. Ecol. Evol.* **5**, 1520 (2021).
3. J. Meckling, P. Y. Lipscy, J. J. Finnegan, F. Metz, *Science* **378**, 31 (2022).
4. X. Chen, Z. Li, K. P. Gallagher, D. L. Mauzerall, *Appl. Energy* **300**, 117318 (2021).
5. Global Energy Monitor, Global coal plant tracker (2022); https://globalenergymonitor.org/projects/global-coal-plant-tracker/.
6. "Dear China, it's time to talk," *Diálogo Chino*, 20 July 2020; https://dialogochino.net/en/extractive-industries/36527-dear-china-its-time-to-talk/.
7. X. Zhou, C. Wilson, B. Caldecott, "The energy transition and changing financing costs" (Univ. of Oxford, 2021); https://www.smithschool.ox.ac.uk/sites/default/files/2022-02/The-energy-transition-and-changing-financing-costs.pdf.
8. Xinhua, "China's central bank steps up support for clean, efficient coal use" (The State Council, The People's Republic of China, 2022); http://english.www.gov.cn/statecouncil/ministries/202205/05/content_WS627305b2c6d02e533532a3d2.html.
9. World Bank, "State and trends of carbon pricing 2022" (World Bank, 2022); http://hdl.handle.net/10986/37455.
10. National Development and Reform Commission (NDRC), "Notice on further improving the coal market price formation mechanism" (2022); https://www.ndrc.gov.cn/xxgk/zcfb/tz/202202/t20220225_1317003_ext.html.
11. C. Nedopil *et al.*, "Green Development Guidance for BRI Projects baseline study report" (BRIGC, 2020); http://en.brigc.net/Reports/Report_Download/202012/P020201201717466274510.pdf.
12. BRIGC, "China to stop new coal abroad for greener BRI," *Belt and Road Portal*, 24 September 2021; https://eng.yidaiyilu.gov.cn/info/iList.jsp?tm_id=139&cat_id=10058&info_id=188312.
13. W. Zhang *et al.*, *Clim. Policy* **23**, 11 (2021).
14. I. Suarez, "BRIEFING: 12.8 GW of Chinese overseas coal project cancelled, but 57 GW could still go ahead" (Centre for Research on Energy and Clean Air, 2022); https://energyandcleanair.org/2022-chinese-overseas-coal.
15. D. Normile, "'A rather beautiful concept': Plan aims to replace Asia's coal plants with renewable energy," *Science*, 2 November 2021; https://www.science.org/content/article/rather-beautiful-concept-plan-aims-replace-asia-s-coal-plants-renewable-energy.

ACKNOWLEDGMENTS

The author thanks D. de Boer, F. Champenois, C. Chang, K. Gallagher, H. Gao, C. Han Springer, and E. Wang for feedback on the drafts and data evaluation support, as well as unnamed Chinese and international policy-makers, investors, and developers for their insights. The author was an independent international researcher of the Green Development Guidance published by BRIGC and an independent international researcher for CCICED studies. All views expressed are the author's and do not represent any official position. Data and materials used in the analysis are available to any researcher for the purposes of reproducing or extending the analysis.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf0126

10.1126/science.adf0126

PERSPECTIVES

CANCER

A common cancer at an uncommon age

The etiology of early-onset colorectal cancer needs to be understood to tackle rising incidence

By **Marios Giannakis**[1,2] *and* **Kimmie Ng**[1]

Early-onset colorectal cancer (EOCRC), also called young-onset colorectal cancer, is defined as CRC diagnosed in individuals aged less than 50 years. EOCRC is increasing globally and anticipated to become the leading cause of cancer death in individuals aged 20 to 49 in the US by 2030 (*1*). Since the 1990s, the age-adjusted incidence of EOCRC has risen at an alarming rate of 2 to 4% per year in many countries, with even sharper increases in individuals younger than 30 years (*1*). This is despite a reduction in overall CRC incidence that is likely attributable to improved screening and prevention in older individuals. The exact reasons and pathophysiology behind the rising incidence of EOCRC remain unknown. Currently, only limited studies exist and they have focused on single aspects of EOCRC etiology. A multidisciplinary path forward is needed to expand the understanding of this increasingly prevalent problem.

EOCRCs exhibit a distinct clinical presentation with a predilection for the left side of the colon and rectum, and they most commonly present with symptoms such as abdominal pain and rectal bleeding (*1*). Patients with EOCRC are often diagnosed with more advanced stage disease, which could result from lack of screening that can detect early lesions, but also raises the question of a more aggressive biology. Indeed, patients with metastatic EOCRC do not have superior survival compared to those with metastatic later-onset CRC (LOCRC) despite fewer comorbidities, better functional status, more frequent utilization of surgery and radiotherapy, higher chemotherapy dose intensity, and fewer adverse events from treatment (*2*).

The increase in EOCRC incidence reflects a "birth cohort" effect, in which the increased risk is carried through generations owing to temporal changes in environmental risk factors that disproportionately affect those born in recent decades compared to those born earlier (*1*). Patients with EOCRC have a higher relative prevalence of inherited predisposition to cancer, with Lynch syndrome being the most common cause. This condition is characterized by deficiency in the DNA mismatch repair pathway that results in high levels of microsatellite instability with an increased number of mutations, which predisposes to CRC and other types of cancer. Although underdiagnosis of Lynch syndrome cases may be a potential contributing factor, this and other high-penetrance pathogenic germline variants do not explain the observed rise of EOCRC. Polygenic risk scores (PRSs) have been devised to select younger individuals for tailored CRC screening, and their performance improves when integrated with environmental risk scores. However, the variants incorporated into these PRSs are obtained from genome-wide association study (GWAS) loci of overall CRC risk (across all ages). Therefore, large GWASs dedicated to EOCRC, as well as analyses of gene-environment interactions, are needed to further refine any genetic contribution that is specific to young-onset presentation.

Several environmental risk factors, including early-life exposures, have been shown or proposed to contribute to the rising incidence of EOCRC. Obesity and other conditions related to metabolic syndrome have globally increased in recent decades, and these factors are also associated with CRC risk. Among participants in the Nurses' Health Study 2, a prospective cohort of healthy nurses aged 25 to 42 years at enrollment who have been followed with validated diet and lifestyle questionnaires over decades, obesity in adolescence and adulthood (*3*) and prolonged sedentary behavior (*4*) were found to be associated with a higher risk of EOCRC. In another study, patients with metabolic conditions such as hypertension, hyperlipidemia, hyperglycemia, and type 2 diabetes mellitus were also more likely to develop EOCRC (*5*). Dietary factors that are now increasingly consumed throughout childhood and adolescence, such as sugar-sweetened beverages, red and processed meat, and Western-pattern diets, have also been implicated. Potentially due to the Westernization of diets and lifestyle, EOCRC incidence is now also rising in low- and middle-income countries. In addition, a host of other potential risk factors have been proposed to be related to EOCRC, including increased use of antibiotics, more ubiquitous environmental toxins, and higher rates of Cesarean sections and other surgical procedures (*1*).

Unfortunately, observational studies only scratch the surface of our understanding of the biology of EOCRC, and efforts to deconvolute the likely multifactorial etiology of EOCRC are hampered by several challenges. Robust epidemiologic studies with validated, repeated, and prospectively collected dietary and lifestyle data across the life continuum (the "exposome") are critically important to accurately measure exposures, their potential confounders, and the time window and cancer latency of the culprit risk factors. Yet, such studies are rare in adults and largely absent among children or adolescents, possibly owing to the complexity and cost of conducting and maintaining such cohorts. In addition, these prospective cohorts should ideally include matched collection of serial biospecimens, such as blood, tissue, and stool across time, to enable a detailed investigation into the underlying mechanisms elicited by environmental exposures in the tumor, tumor microenvironment (TME), and gut microbiota. Studying the interactions between the exposome, tumor-TME, and host will be fundamental to uncovering the root causes of the rise in EOCRC.

A few studies have attempted to profile the somatic mutational landscape of EOCRCs using next-generation sequencing panels. Such analyses found that somatic mutations in genes encoding members of signaling pathways that are known to be oncogenic drivers were differentially present: adenomatous polyposis coli (*APC*) and *BRAF* were less frequently mutated in EOCRC (*6*, *7*), whereas *TP53* and β-catenin (*CTNNB1*) mutations occurred more often in EOCRC (*7*). However, a subsequent study found no differences between EOCRC and LOCRC somatic mutations when sidedness of the tumor was taken into account (*8*). Such studies highlight some of the

[1]Young-Onset Colorectal Cancer Center, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA. [2]Broad Institute of MIT and Harvard, Cambridge, MA, USA. Email: marios_giannakis@dfci.harvard.edu; kimmie_ng@dfci.harvard.edu

challenges and caveats in defining the molecular landscape of EOCRC, which include adjusting for confounding clinical and pathologic variables. Specifically, sidedness of the tumor, as well as frequency of high microsatellite instability, both of which differ between EOCRC and LOCRC, affect the somatic mutational landscape. Left-sided tumors, compared to their right-sided counterparts, have different embryologic origins and are exposed to factors that vary along the gut, both of which may explain the different mutational profiles observed in various regions of the colon.

numbers of EOCRCs and integration with epidemiologic data could uncover mutagenic processes that contribute to tumorigenesis across the age continuum and strengthen support for causation. Beyond genomics, hypomethylation of long interspersed nuclear element 1 (LINE-1) transposable elements is an epigenetic feature that is more common with decreasing age of CRC diagnosis (*1*). Single-cell RNA sequencing (scRNA-seq) efforts in CRC (predominantly from older patients) deconvoluted 88 cell subsets and 204 associated gene expression programs (*11*), offer-

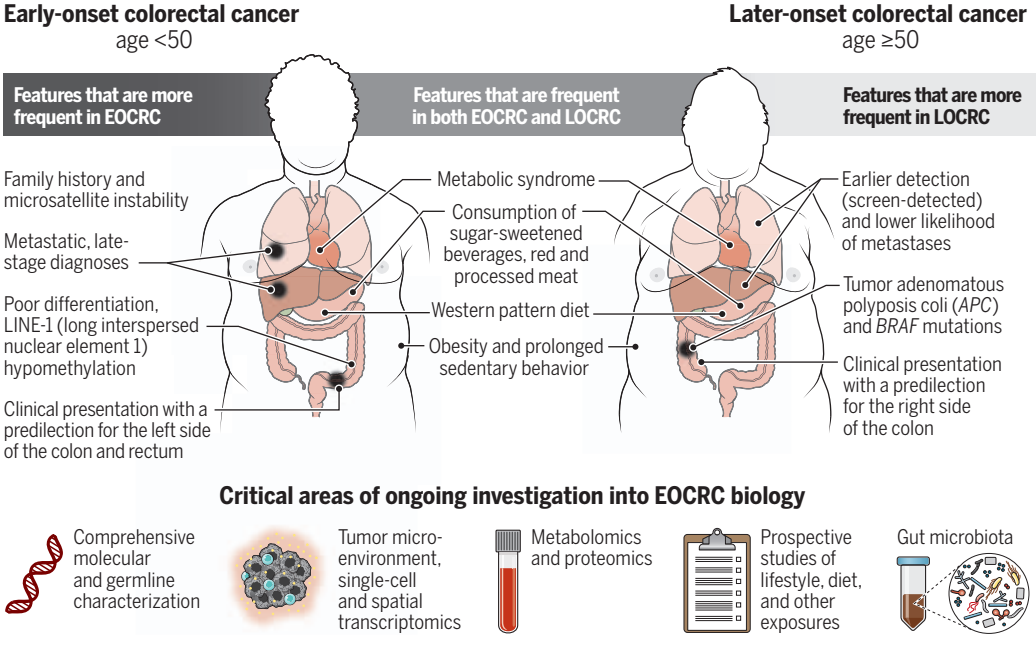
demonstrated how lifestyle factors can affect the TME of incident cancers, such as the association of smoking with risk of CRCs with low T cell infiltrates (*14*). The scRNA-seq efforts have revealed coordinated spatially organized interaction hubs of malignant and nonmalignant cells in primary CRCs that differed among mismatch repair–proficient and –deficient CRCs and highlighted the role of immune and stromal cells in malignant progression (*12*). These datasets can be leveraged for future mechanistic investigation and drug target discovery. Computational approaches are also being further refined and developed to comprehensively describe the spatial organization and functional interactions among individual cells in the CRC TME.

Given the compelling body of evidence supporting a role of the gut microbiota in CRC pathogenesis and progression, including species such as *Fusobacterium nucleatum*, *Bacteroidetes fragilis,* and *pks*$^{+}$ *E. coli,* studies of EOCRC should also profile the tumor and stool microbiomes of EOCRC patients. For example, differences have been reported in fecal microbial composition, diversity, and function in EOCRC compared to LOCRC and age-matched healthy controls in China (*15*). These findings need to be confirmed in larger and more diverse populations, but they reveal the potential of the microbiome in EOCRC early detection and prognostic assessment. In addition, experimental models of EOCRC, such as patient-derived organoid cultures and animal models, will need to incorporate elements of the CRC TME, such as the microbiota. However, standardizing protocols that minimize variability and systematically incorporating microbiome specimens into prospective cohort and trial designs pose substantial logistical challenges.

The path forward to combat the rise of EOCRC is neither short nor straightforward. The recent recommendations by the American Cancer Society and US Preventive Services Task Force to start CRC screening for the average population at age 45 (versus the prior recommended age of 50) are a first step in recognizing this epidemic. However, the urgency of navigating this path, which goes beyond unidimensional perspectives

Colorectal cancer in younger people

Similar factors increase the risk of early-onset colorectal cancer (EOCRC) and later-onset colorectal cancer (LOCRC), such as a sedentary lifestyle, obesity, and metabolic syndrome, but there are also important differences. EOCRC predominantly occurs on the left side of the colon and the rectum, whereas LOCRC arises more commonly on the right side of the colon. EOCRC is also more poorly differentiated and often metastatic at diagnosis. Research is urgently needed to understand the increasing incidence of EOCRC and its pathophysiology to better detect and treat patients.



Moreover, although targeted next-generation sequencing panels are useful in clinical practice for identifying actionable alterations in CRC, they cannot characterize the full spectrum of molecular alterations in young-onset tumors.

However, whole-genome sequencing allows for the identification of noncoding elements, neoantigens, and mutational signatures (distinct patterns of mutations) that can be linked to specific macro- or microenvironmental mutagens and CRC pathogenesis, such as high prediagnosis consumption of red meat (*9*) and genotoxic *Escherichia coli* that express the polyketide synthase (*pks*) island (*10*). The deconvolution of mutational signatures from large

ing an unprecedented view into abnormal CRC cell states. Single-cell transcriptional and epigenetic profiling of hereditary and sporadic premalignant lesions is similarly revealing changes along the continuum of tumorigenesis (*12*). Application of these approaches in EOCRCs could prove valuable in revealing contrasting pathophysiology compared to LOCRC that may point to the etiology of EOCRC (see the figure).

The TME in EOCRCs also warrants investigation, especially because it can be readily molded by environmental risk factors. The immune contexture—the type, density, and location of immune cells in the colorectal TME—has prognostic importance (*13*), and molecular epidemiologic studies have

GRAPHIC: K. HOLOSKI/*SCIENCE*

and considers the multifactorial nature of EOCRC, is paramount, particularly for the youngest patients who do not meet the recommended screening age.

What steps can be taken to map out this path? The emergence of specialized centers of excellence that are focused on patients with EOCRC will establish a model of comprehensive clinical care for this population, as well as enabling multidisciplinary research. Prospective cohort studies of healthy individuals and patients with EOCRC are needed. These should be accompanied by serial measurements of the exposome paired with biospecimen collections. Moreover, the pace of progress needs to be accelerated by forming global collaborations to facilitate patient and biospecimen accrual, and by implementing innovative models of patient recruitment such as the Count Me In Colorectal Cancer Project (https://joincountmein.org/colorectal), which directly partners with patients in the US and Canada and makes all data available for research. Effort is needed to ensure that diverse populations are included in studies of EOCRC, particularly underrepresented minorities who are disproportionately burdened by EOCRC, as evidenced by the higher mortality of non-Hispanic Black EOCRC patients compared to non-Hispanic white patients (*1*). Consideration also needs to be given to implementation of screening in younger age groups, as well as earlier detection using blood-based biomarkers. Although each of these steps requires commitment and perseverance, it is the growing numbers of young patients bravely battling this disease that will be the compass that keeps us on the path towards better understanding, preventing, and treating EOCRC. ■

REFERENCES AND NOTES

1. N. Akimoto *et al.*, *Nat. Rev. Clin. Oncol.* **18**, 230 (2021).
2. M. Lipsyc-Sharf *et al.*, *J. Natl. Cancer Inst.* **114**, 427 (2022).
3. P. H. Liu *et al.*, *JAMA Oncol.* **5**, 37 (2019).
4. H. Chen *et al.*, *Gut* **70**, 1147 (2021).
5. L. H. Nguyen *et al.*, *JNCI Cancer Spectr.* **2**, pky073 (2018).
6. A. N. Willauer *et al.*, *Cancer* **125**, 2002 (2019).
7. C. H. Lieu *et al.*, *Clin. Cancer Res.* **25**, 5852 (2019).
8. A. Cercek *et al.*, *J. Natl. Cancer Inst.* **113**, 1683 (2021).
9. C. Gurjao *et al.*, *Cancer Discov.* **11**, 2446 (2021).
10. C. Pleguezuelos-Manzano *et al.*, *Nature* **580**, 269 (2020).
11. K. Pelka *et al.*, *Cell* **184**, 4734 (2021).
12. W. R. Becker *et al.*, *Nat. Genet.* **54**, 985 (2022).
13. J. Galon *et al.*, *Science* **313**, 1960 (2006).
14. T. Hamada *et al.*, *J. Natl. Cancer Inst.* **111**, 42 (2019).
15. Y. Yang *et al.*, *Nat. Commun.* **12**, 6757 (2021).

ACKNOWLEDGMENTS

The authors thank B. Cahill for assistance with the figure. M.G. has received research funding from Janssen and Servier. K.N. has received institutional research funding from Pharmavite, Evergrande Group, Janssen, and Revolution Medicines; and advisory or consulting fees from Bayer, GlaxoSmithKline, and Pfizer.

10.1126/science.ade7114

NEUROSCIENCE

A cryptic clue to neurodegeneration?

Antisense oligonucleotides rescue cryptic RNA splicing and neuron regeneration

By **Niamh O'Brien**[1,2] *and* **Sarah Mizielinska**[1,2]

The specialized function of neurons relies heavily on alternative splicing of RNA, and dysregulation of this process is implicated across the neurodegenerative disease spectrum. Advances in RNA sequencing have enabled the discovery of aberrations in splice recognition sites. This includes those that lead to misprocessing of precursor mRNAs (pre-mRNAs), resulting in the inclusion of so-called "cryptic exons" and the production of truncated mRNAs (and sometimes corresponding truncated proteins). On page 1140 of this issue, Baughn *et al.* (*1*) report the mechanism of cryptic exon missplicing of *STMN2* RNA (encoding the protein stathmin-2), which occurs when a protein associated with neurodegenerative disease—TAR DNA-binding protein of 43 kDa (TDP-43)—is absent from the nucleus. Treatment with antisense oligonucleotides (ASOs) restored normal splicing and stathmin-2 levels in cultured motor neurons and in a mouse model. These findings point to the possibility of therapeutically targeting cryptic exons to prevent RNA missplicing and associated disease.

TDP-43 proteinopathy is a term used to describe neurodegenerative diseases that exhibit aggregation of the DNA- and RNA-binding protein TDP-43 irrespective of disease type. Initially, this was mostly of interest in the study of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD), in which it is a dominant pathology, and mutations in TDP-43 are genetically associated with these diseases (*2*). However, TDP-43 proteinopathy has recently been discovered in other neurodegenerative disorders, notably in Alzheimer's disease (*3*). TDP-43 can be present in variable forms of aggregates in neurons and glia, with post-translational modifications and C-terminal truncation (*4*). Cytoplasmic aggregation is frequently concurrent with its depletion from the nucleus, where RNA splicing occurs.

The roles of TDP-43 in RNA homeostasis are diverse, including the regulation of RNA transcription, alternative splicing, and transport. Depletion of TDP-43 from the nucleus leads to altered amounts of hundreds of RNA transcripts, some of which exhibit missplicing onto cryptic exons (*5*). Cryptic exons are a class of exons that are found within noncoding intronic regions of pre-mRNA but can be spliced into mature RNA, often leading to a premature stop codon. The absence of TDP-43 from the nucleus of motor neurons results in decreased expression of stathmin-2, which has roles in axonal regeneration, microtubule stability, and lysosome trafficking (*6*). This reduction is the result of missplicing onto a cryptic exon and production of a nonfunctional mRNA that is degraded (*6*, *7*). Notably, *STMN2* RNA cryptic splicing is found in genetic and sporadic forms of ALS and FTD and is specific to TDP-43 proteinopathy (*6*–*8*), suggesting its potential as a therapeutic target or biomarker.

Baughn *et al.* mapped the TDP-43 binding site on *STMN2* pre-mRNA in mice to a stretch of three closely spaced guanine (G)–uracil (U) hexamers in intron 1 of the gene. Deleting this sequence led to cryptic exon splicing, demonstrating that TDP-43 binding is essential for maturation of *STMN2* pre-mRNA and production of functional stathmin-2. Replacement of the GU hexamers with another sequence and associated binding protein prevented cryptic splicing, suggesting that TDP-43 acts through steric hindrance of splicing factors. This effect was specific to the cryptic exon binding site and not a nearby cryptic polyadenylation site, suggesting alternative polyadenylation sites downstream.

Defining the molecular interaction of *STMN2* pre-mRNA and TDP-43 enabled subsequent targeting of cryptic exon splicing with ASOs. This is a promising therapeutic approach for modulating gene expression because these short single-stranded DNA molecules can selectively target RNA transcripts. ASOs have been approved for clinical use in diseases such as spinal muscular atrophy and Duchenne muscular dystrophy. Baughn *et al.* designed a series of ASOs to target the *STMN2*

[1]UK Dementia Research Institute at King's College London, London, UK. [2]Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College London, Maurice Wohl Clinical Neuroscience Institute, Camberwell, London, UK. Email: sarah.mizielinska@kcl.ac.uk

cryptic exon. These oligonucleotides restored stathmin-2 expression in motor neurons (derived from induced human pluripotent stem cells) with reduced TDP-43. Functionally, this treatment restored axon regeneration (after axotomy) and improved stathmin-2–dependent lysosome trafficking and synaptic active zone organization.

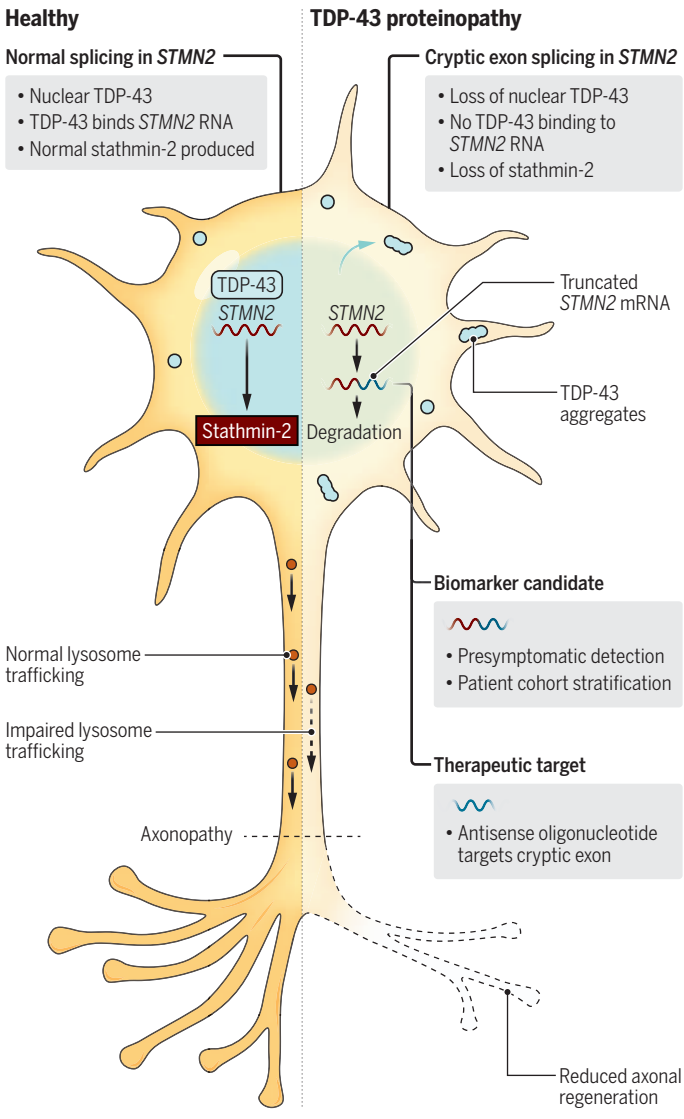
Restoration of correct *STMN2* pre-mRNA processing was also shown in vivo. Because mice do not have the *STMN2* cryptic exon sequence, mice were genetically engineered to bear a partially humanized intron 1 (mouse sequence replaced with equivalent human sequence) but with the TDP-43 binding site (the GU hexamer) deleted. In heterozygous animals (one normal and one mutated allele), the inability of TDP-43 to bind to the corresponding RNA transcript resulted in cryptic splicing and a reduction in stathmin-2 expression, despite normal amounts of nuclear TDP-43. Despite this loss of stathmin-2, heterozygous mice developed normally. However, genetically engineered mice in which both alleles were altered showed substantially decreased survival, demonstrating a dose-dependent requirement for stathmin-2. ASOs targeting the cryptic exon partially ameliorated these effects. Conversely, insertion of the cryptic exon into both alleles of mice harboring the ALS-associated TDP-43 Q331K mutation (which develop age-related motor neuron disease) neither exacerbated the disease phenotype nor resulted in *Stmn2* RNA cryptic splicing. This confirms the requirement for the loss of nuclear TDP-43 (which is not present in the mouse model) for cryptic splicing alterations and highlights the challenges of finding appropriate preclinical models for the study of TDP-43 proteinopathy.

The study by Baughn *et al.* demonstrates that cryptic exons can be targeted with ASOs that can restore physiological splicing and functional protein production (see the figure). The importance of stathmin-2 in motor neurons was strengthened by a recent study that demonstrated distal motor neuropathy in mice that were genetically engineered to lack the protein (*9*). However, whether the restoration of stathmin-2 is beneficial in neurodegeneration and thus clinical disease across TDP-43 proteinopathies is yet to be determined. Recently, genetic association of a repeat expansion in *STMN2* has been identified in sporadic ALS (*10*), supporting a direct role of stathmin-2 in ALS. Because *STMN2* is one of many TDP-43 targets, a wider approach to gene expression changes may also be required for further functional improvement; other targets include *UNC13A* (unc-13 homolog A), which encodes a synaptic protein associated with FTD and ALS (*11*). Its corresponding RNA is also cryptically spliced, and its expression decreases upon loss of TDP-43 from the nucleus (*12*, *13*). Targeting TDP-43 could have broad effects across the neurodegenerative disease spectrum, and efforts are ongoing in this area, but an advantage of targeting downstream effectors, particularly disease-specific effects such as cryptic splicing, is that therapeutics will only target vulnerable neurons where pathology is present.

An alternative and exciting avenue from the discovery of cryptic exon splicing is the potential of de novo truncated mRNAs and protein as biomarkers. Biomarkers enable the monitoring of target engagement by a therapeutic agent and, consequently, improvements in preclinical and clinical studies. They can also be used to stratify patient cohorts and detect presymptomatic individuals, which is crucial for effective clinical trials. This would benefit the clinical study of TDP-43 proteinopathies, which have diverse clinical presentation and are predominantly sporadic. Targeting TDP-43 as a biomarker has been challenging because fluid biomarkers often display a lack of sensitivity to detection, and effective imaging with positron emission tomography has yet to be developed (*14*). Two preliminary studies point to the detection of de novo proteins from cryptic splicing in the cerebrospinal fluid of presymptomatic individuals with ALS or FTD (*15*, *16*). Thus, cryptic splice products may fill a critical gap. ■

Cryptic exon splicing

TDP-43 (TAR DNA-binding protein of 43 kDa) binds to *STMN2* precursor mRNA that encodes stathmin-2, ensuring normal splicing and protein expression (left). Loss of nuclear TDP-43 results in the inclusion of a cryptic exon in mature *STMN2* mRNA that is then degraded (right). Cryptic splicing of *STMN2* RNA has potential as a biomarker for TDP-43 proteinopathies and as a target for antisense oligonucleotide therapies.



GRAPHIC: A. MASTIN/*SCIENCE*

REFERENCES AND NOTES

1. M. W. Baughn *et al.*, *Science* **379**, 1140 (2023).
2. M. Neumann, *Biochem. Biophys. Res. Commun.* **10**, 232 (2006).
3. K. A. Josephs *et al.*, *Acta Neuropathol.* **127**, 811 (2014).
4. E. L. Sternburg *et al.*, *Trends Biosci.* **47**, 6 (2022).
5. J. P. Ling *et al.*, *Science* **349**, 650 (2015).
6. J. R. Klim *et al.*, *Nat. Neurosci.* **22**, 167 (2019).
7. Z. Melamed *et al.*, *Nat. Neurosci.* **22**, 180 (2019).
8. M. Prudencio *et al.*, *J. Clin. Invest.* **130**, 6080 (2020).
9. K. L. Krus *et al.*, *Cell Rep.* **39**, 111001 (2022).
10. F. Theunissen *et al.*, *Front. Aging Neurosci.* **13**, 658226 (2021).
11. M. A. van Es *et al.*, *Nat. Genet.* **41**, 1083 (2009).
12. A. L. Brown *et al.*, *Nature* **603**, 131 (2022).
13. X. R. Ma *et al.*, *Nature* **603**, 124 (2022).
14. P. Steinacker *et al.*, *Mol. Cell. Neuro.* **97** (2019)
15. S. Seddighi *et al.*, bioRxiv 10.1101/2023.01.23.525149 (2023).
16. K. E. Irwin *et al.*, bioRxiv 10.1101/2023.01.23.525202 (2023).

ACKNOWLEDGMENTS

N.O. and S.M. are supported by MND Association Grant Ruepp/Apr19/872-791.

10.1126/science.adg8501

IMMUNOLOGY

Autoimmunity to the modified self

Protein posttranslational modifications can break tolerance to the self-proteome

By **Laura Santambrogio**[1,2,3]

Hundreds of protein posttranslational modifications (PTMs) have been mapped, greatly expanding the complexity of the cellular proteome and substantially diversifying its functions (*1*). By changing the protein primary structure, PTMs may change protein function. Additionally, by modifying the self-proteome, PTMs pose a danger for the development of autoimmune diseases because they change the protein "self" sequence. On page 1104 of this issue, Zhai *et al.* (*2*) describe how carboxyethylation of integrin αIIb (ITGA2B) is involved in the pathogenesis of the autoimmune disorder, ankylosing spondylitis.

PTMs such as phosphorylation, ubiquitination, acetylation, and glycosylation are crucial in regulating cellular processes and molecular functions. Conversely, other PTMs such as glycation, carbonylation, oxidation, lipoxidation, and citrullination are mostly observed during pathological conditions that are associated with acute and chronic inflammation. During these pathological conditions, a cellular redox imbalance generates an increased amount of reactive oxygen species and reactive nitrogen species, which oxidize biomolecules (*3*). Similarly, in metabolic conditions such as type 2 diabetes mellitus, increased glycemia (blood sugar concentration) induces protein glycation, glycoxidation, and lipoxidation (*4*). PTM-modified cytosolic proteins are disposed of by the proteasome, or by the endolysosomal compartment through autophagy (*5*). Extracellular PTM-modified proteins enter the endocytic pathway through phagocytosis. The proteasome and the endolysosomal compartment enzymatically digest the modified proteins and generate PTM-modified peptides for presentation on major histocompatibility complex (MHC) class I and MHC class II molecules (*5*).

Through their T cell receptors (TCRs), T cells recognize self and nonself (pathogen) proteins through their presentation as processed peptides by MHC class I and class II complexes expressed by all cells (MHC class I) or antigen-presenting cells (MHC class II). MHC molecules are the most polymorphic proteins in humans, encoded by hundreds of different human leukocyte antigen (HLA) alleles. Their polymorphism is mostly clustered in the peptide binding groove that drives distinctive peptide selection and binding specificities. The selectivity of each MHC variant for different sets of self and nonself-peptides shapes the differences in adaptive immune responses across the population and is an important driver in autoimmune disease.

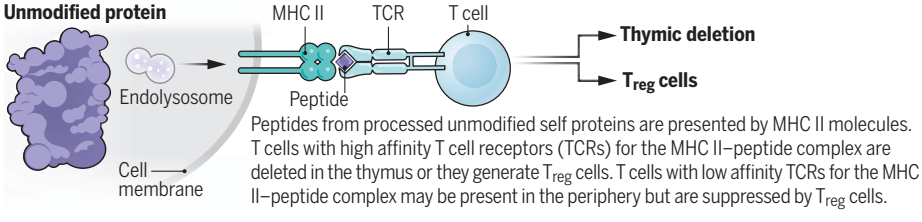
Upon presentation by MHC class I or class II molecules, PTM-modified peptides can change the interaction with the cognate TCR and thus the outcome of the immune response. For example, PTMs can decrease or increase the affinity of the peptide for the MHC binding groove, thus affecting the overall composition of the MHC peptidome presented by antigen-presenting cells to T cells (*4*, *6*, *7*). PTMs may also change the structure of the peptide amino acids involved in TCR binding, selecting a different T cell population compared with the one that would recognize the unmodified peptide. PTM-modified peptides can also be recognized by the immune system as nonself, thus triggering an autoimmune response (*2*, *4*, *6*, *7*). For example, immune responses to oxidized peptides have been reported in atherosclerosis and cardiovascular diseases in which oxidized phospholipids and malondialdehyde-modified peptides are recognized by autoreactive T and B cells (*8*). Similarly, citrullinated modifications of vimentin and cartilage-related proteins as well as carbamylated proteins are targeted by autoreactive T cells in rheumatoid arthritis (*9*). Further, acetylated histone peptides are targeted in systemic lupus erythematosus by autoantibodies whose titer and activity correlate with disease severity (*10*).

Zhai *et al.* describe cysteine carboxyethylation of ITGA2B by cystathionine β synthase (CBS). This reaction modifies cysteine residues in ITGA2B in patients with ankylosing spondylitis, thus increasing the risk of nonself immune recognition of PTM-modified ITGA2B. In a subset of these patients, who share the HLA-DRB1*04 haplotype, the carboxyethylated-Cys^{96} ITGA2B peptide has a high affinity for this MHC class II haplotype and thus activates antigen-specific autoreactive T cells (see the figure). Similarly, increased B and T cell reactivity was observed in HLA-DR4 mice immunized with the same peptide. The T and B cell autoreactive response to the modified peptide is sufficient to induce extracellular matrix inflammation and damage leading to the pathogenesis of ankylosing spondylitis in mice.

The pathogenic autoreactivity reported by Zhai *et al.* raises questions about the maintenance of self-tolerance to proteins and peptides after PTMs. During thymocyte maturation, several mechanisms ensure that the maximum number of self-antigens are

Breaking tolerance with posttranslational modifications

Proteins are presented on cell-surface major histocompatibility complex (MHC) class I and class II molecules for recognition by adaptive immune cells. Tolerance to self proteins is achieved through deletion of reactive T and B cells and suppression by regulatory T cells (T_{reg} cells). But if a self protein contains unusual posttranslational modifications (PTMs), it could be presented by MHC molecules and recognized by T cells, inducing autoimmunity.



Peptides from processed unmodified self proteins are presented by MHC II molecules. T cells with high affinity T cell receptors (TCRs) for the MHC II–peptide complex are deleted in the thymus or they generate T_{reg} cells. T cells with low affinity TCRs for the MHC II–peptide complex may be present in the periphery but are suppressed by T_{reg} cells.

T cells specific for MHC II peptides from proteins that are carboxyethylated can escape thymic deletion and may clonally expand into pathogenic effector T cells when the PTM-modified peptide is presented by a high-affinity MHC II haplotype, thereby inducing autoimmunity.

GRAPHIC: N. CARY/*SCIENCE*

[1]Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY, USA. [2]Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA. [3]Department of Radiation Oncology, Weill Cornell Medicine, New York, NY, USA.
Email: las4011@med.cornell.edu

presented during thymic selection. These mechanisms include presentation of tissue-restricted antigens by medullary thymic epithelial cells through autoimmune regulator (AIRE)–mediated and Fez family zinc finger protein 2 (FEZF2)–mediated transcriptional regulation (*11*), as well as presentation of peripherally processed peptides that are transported to the thymus by migratory dendritic cells (*12*). Among these self-antigens, it is conceivable that many peptides will contain physiological PTMs. However, PTMs that are mostly observed in pathologic conditions may not be present in the MHC-peptide repertoire presented to instruct maturing T cells toward tolerance, and thus, they potentially could induce autoimmunity.

To avoid autoimmunity, thymus-derived regulatory T cells (T_{reg} cells) directly control ~30% of autoreactive conventional T cells from converting into pathogenic effectors (*13*). Although T_{reg} cell activation is MHC-peptide specific, once activated, T_{reg} cells can suppress conventional T cells specific for a different MHC-peptide complex, through secretion of anti-inflammatory cytokines and overall bystander suppression. Additionally, the circulating T cell repertoire includes peripheral T_{reg} cells, generated from the differentiation of conventional naïve T cells that are repetitively stimulated with suboptimal antigen concentration or by peptides derived from the commensal microbiota (*14*). These thymic or peripherally derived T_{reg} cells may control T cells that are reactive to PTM-modified self-peptides that arise during sterile and pathogen-induced inflammatory conditions and during metabolic disease. However, in the presence of an MHC haplotype with high affinity for the PTM-modified peptide—for example, HLA-DRB1*04—and a T cell repertoire that strongly favors PTM-peptide recognition, autoreactive T cells can escape T_{reg} cell–mediated suppression and clonally expand into pathogenic effector T cells. Given the large number of PTMs mapped during inflammatory and metabolic conditions, many more examples of recognition of PTM-modified self proteins in autoimmunity are likely to be identified. ■

REFERENCES AND NOTES

1. S. Ramazi, J. Zahiri, Database 2021, baab012 (2021).
2. Y. Zhai *et al.*, *Science* **379**, 1104 (2023).
3. E. S. Cannizzo *et al.*, *J. Proteomics* **74**, 2313 (2011).
4. C. C. Clement *et al.*, *Immunity* **54**, 721 (2021).
5. E. S. Cannizzo *et al.*, *Cell Rep.* **2**, 136 (2012).
6. C. C. Clement *et al.*, *Sci. Immunol.* **7**, eabl3795 (2022).
7. C. C. Clement *et al.*, *JCI Insight* **1**, e85633 (2016).
8. A. Taleb *et al.*, *Biomarkers Med.* **5**, 673 (2011).
9. M. F. Konig *et al.*, *Ann. Rheum. Dis.* **75**, 2022 (2016).
10. J. Dieker *et al.*, *PLOS ONE* **11**, e0165373 (2016).
11. R. Benlaribi *et al.*, *Inflamm. Regen.* **42**, 28 (2022).
12. R. Bonasio *et al.*, *Nat. Immunol.* **7**, 1092 (2006).
13. A. Cebula *et al.*, *Nat. Commun.* **10**, 4882 (2019).
14. S. Z. Josefowicz *et al.*, *Annu. Rev. Immunol.* **30**, 531 (2012).

10.1126/science.adg3925

MOLECULAR BIOLOGY

A massive machine regulates cell death

Structural analysis reveals how the decision to induce apoptotic cell death is regulated

By **Peter D. Mace** *and* **Catherine L. Day**

Cells are bombarded with signals about their environment, which they integrate to decide an appropriate response. Danger signals elicit the most drastic of these decisions—is the situation salvageable, or should the cell be sacrificed? When danger signals predominate, one of several pathways induces cell death, with apoptosis being the most common. The key executioners of apoptosis are proteases called caspases; when caspases are activated, apoptosis becomes irreversible. Caspase activation is tightly controlled by regulatory molecules, including the inhibitor of apoptosis (IAP) proteins. The largest and most diverse member of the IAP family, baculoviral IAP repeat–containing protein 6 (BIRC6), has remained an enigma. On pages 1105, 1112, and 1117 of this issue, Hunkeler *et al.* (*1*), Dietz *et al.* (*2*), and Ehrmann *et al.* (*3*), respectively, reveal the molecular basis of how BIRC6 controls cell fate, which may ultimately inform the development of new anticancer drugs that induce cell death.

Caspases are expressed as inactive zymogens, with dimerization and cleavage of the N-terminal prodomain required for full activity (*4*). In unstressed cells, low-level caspase activity is kept in check by IAPs. All IAPs contain one or more baculoviral IAP repeat (BIR) domains, which are required for caspase inhibition. Although there are subtle differences among IAP family members, the ability of a shallow groove on the BIR domain to bind a four–amino acid motif at the N terminus of active caspases is central to caspase suppression. During apoptosis, the release of second mitochondria–derived activator of caspases (SMAC) from the mitochondrial intermembrane space unleashes caspase protease activity (*5*). The groove on the BIR domain is also central to the ability of SMAC to unleash caspases because SMAC contains a four–amino acid motif that binds to BIR domains more potently than caspases, thereby displacing them from IAPs. Because the goal of many cancer treatments is to trigger caspase activation and cell death, this mechanism of caspase control inspired the development of a range of small-molecule IAP antagonists, known as SMAC mimetics, that mimic the four amino acids that bind the BIR domain to release caspases and trigger apoptosis.

In addition to direct caspase inhibition by the BIR domain, most IAPs have domains that either bind ubiquitin or promote the attachment of ubiquitin to proteins (*6*). For example, the well-studied IAPs [cellular IAP1 (cIAP1), cIAP2, and X chromosome–linked IAP (XIAP)] have a RING domain that confers ubiquitin ligase activity. More than a decade ago, it was discovered that SMAC-mimetic drugs not only act by releasing caspases from XIAP, but, unexpectedly, they also activate the ubiquitin ligase activity of cIAPs (*7*). Thus, the interplay of BIR domain binding and ubiquitin modification is an overriding feature in the ability of IAP proteins to regulate cell death or survival.

BIRC6 is the most unusual member of the IAP family. It is gigantic at 4857 amino acids (approximately 10 times as large as the average human protein), but it only has one BIR domain. Until now, only one other domain had been identified: a ubiquitin-conjugating domain that is not found in other IAPs but defines the E2 ubiquitin-conjugating family of proteins. Ubiquitin-conjugating domains normally function with partner ubiquitin ligase proteins, but BIRC6 appears capable of functioning as a rare hybrid enzyme. Although BIRC6 is an essential protein that is required for mouse embryonic development (*8*, *9*), little has been known about its mechanism of action. Now, using a combination of biochemistry, electron microscopy, and cell biology, several groups have started to uncover how BIRC6 functions (*1*–*3*, *10*).

The size of BIRC6 has historically created issues for structural analysis, which are now surmountable using cryo–electron microscopy. The studies of Hunkeler *et al.*, Dietz *et al.*, and Ehrmann *et al.* present a structural biology tour de force that reveals that BIRC6 is not only large in primary structure but assembles into an even larger head-to-tail

Department of Biochemistry, School of Biomedical Sciences, University of Otago, Dunedin, New Zealand.
Email: peter.mace@otago.ac.nz; catherine.day@otago.ac.nz

dimer. A series of armadillo-like domains mediates dimerization and forms a scaffold on which multiple other domains assemble. The overall effect is a large, U-shaped architecture with a central cavity flanked by key functional domains (see the figure). This includes a large WD40 propeller domain that organizes the N-terminal module and stabilizes the position of the BIR domain so that the key four–amino acid binding groove is oriented toward the central cavity. The twin ubiquitin-conjugating domains appear to hover atop each arm of the U-shaped architecture in a flexible manner.

Multiple structures of BIRC6 in complex with partners show that the central cavity is the key docking site for both substrates and inhibitors alike, including caspase-9, high-temperature requirement protein A2 (HTRA2), and SMAC. This suite of structures reveals how BIRC6 acts as a platform for caspase ubiquitination, is inhibited by SMAC, and acts as a juncture in cellular decision-making between inducing apoptosis or autophagy—a cell survival pathway. They show that BIRC6 is a bona fide ubiquitin ligase for active caspase-3, -7, and -9 and the apoptotic protease HTRA2, and thus, BIRC6 inhibits apoptosis by promoting their proteasomal degradation. Structures with caspase-9 and HTRA2 show that substrates bind within the central cavity of the BIRC6 dimer, interacting with the BIR domains on either side as well as other domains in the central scaffold. The flanking ubiquitin-conjugating domains can then transfer ubiquitin onto substrates in the central cavity. The central cavity acts as an aptly named "zone of ubiquitination" for apoptotic substrates, which are tethered in place by interactions with the BIR domains.

That substrates are commonly bound in the central cavity provides a clear target for SMAC antagonism of BIRC6. Structures of SMAC-bound BIRC6 place the dimeric coiled-coils of SMAC across the central cavity, which itself is tethered by interactions with the BIR domains on either side. Unexpectedly, extra helices of BIRC6 wrap SMAC in an extensive embrace and make SMAC binding near irreversible. Notably, this arrangement excludes substrates from binding to the central cavity or from binding to the BIR domains. These features make SMAC an exquisitely effective BIRC6 antagonist, with a binding surface far more extensive than any other characterized IAP-SMAC complex.

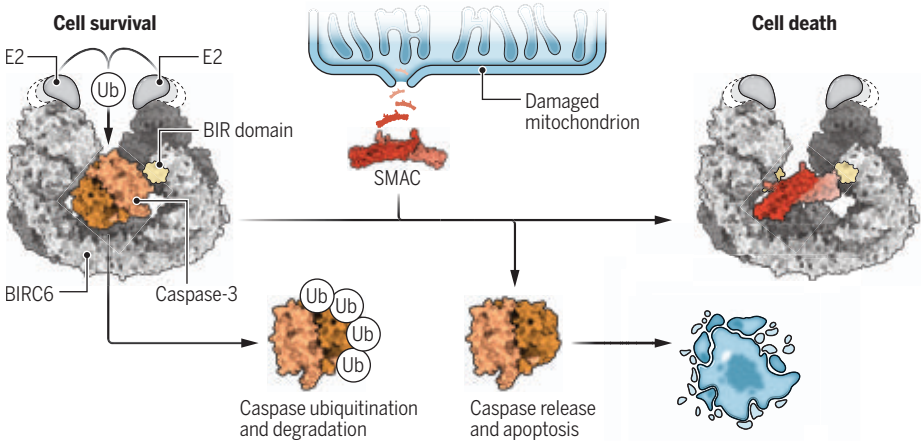
Downstream outcomes of the interplay between BIRC6, substrates, and SMAC likely differ according to cellular circumstance because BIRC6 not only regulates apoptosis but also the cellular salvage pathway of autophagy (*11*). The role of BIRC6 in autophagy is through ubiquitination and degradation of light chain 3 (LC3), which mediates the selective recruitment of cargo to autophagosomes. The study of Ehrmann *et al.* as well as preliminary data (*10*) investigate interactions between BIRC6 and LC3. Although neither visualize binding of LC3 to BIRC6, both show that binding is unlikely to involve the BIR domain. Instead, hydrophobic motifs in central regions of BIRC6 appear critical. Clarifying complete details of BIRC6 substrate binding and the downstream consequences for substrate selection and ultimately cell fate are important for future investigation.

One pertinent corollary of the studies of Hunkeler *et al.*, Dietz *et al.*, and Ehrmann *et al.* is the confirmation that BIRC6-mediated ubiquitination depends on a noncanonical starting point (*11*). Whereas most ubiquitin conjugation in cells is initiated by a canonical E1 enzyme in humans, BIRC6 functions exclusively with a noncanonical E1 [ubiquitin-like modifier-activating enzyme 6 (UBA6)]. There are a limited number of defined partners of UBA6, and Hunkeler *et al.* identified codependency between BIRC6 and UBA6 in cancer (*12*), which may hint at one route to specifically antagonize BIRC6 therapeutically. These studies also unequivocally establish BIRC6 as a hybrid E2-E3 enzyme, like ubiquitin-conjugating enzyme E2 O (UBE2O), which also has substrate binding and ubiquitin transfer activity (*13*, *14*). However, much remains to be understood about the mechanism by which BIRC6 brings about ubiquitin transfer and the types of modifications that it assembles on substrates.

Together, the studies by Hunkeler *et al.*, Dietz *et al.*, and Ehrmann *et al.* advance our understanding of a key protein that regulates cell death. They also clearly establish the importance of the BIR domain for caspase inhibition but show that current SMAC mimetics are not well suited to antagonize BIRC6 because they were optimized for binding to the pocket on the BIR domains of cIAP and XIAP. The stage now seems to be set for the development of SMAC mimetics that specifically target BIRC6. Such compounds would have potential for the direct activation of caspases or could be used to harness the BIRC6 ubiquitination machinery in the emerging field of degradation-based therapeutics (*15*). ■

A life or death decision

Structures of baculoviral inhibitor of apoptosis (IAP) repeat–containing protein 6 (BIRC6) reveal how it regulates apoptosis. When bound to caspases, the BIR domains of the BIRC6 homodimer ensure that substrates are optimally positioned for ubiquitination (Ub) and degradation. However, when second mitochondria–derived activator of caspases (SMAC) is released from damaged mitochondria, it binds more tightly to BIRC6 than caspases do. This near-irreversible inhibition of BIRC6 activates caspase-mediated apoptosis.



GRAPHIC: A. FISHER/*SCIENCE*

REFERENCES AND NOTES

1. M. Hunkeler, C. Y. Jin, E. S. Fischer, *Science* **379**, 1105 (2023).
2. L. Dietz *et al.*, *Science* **379**, 1112 (2023).
3. J. F. Ehrmann *et al.*, *Science* **379**, 1117 (2023).
4. S. Kesavardhana, R. K. S. Malireddi, T.-D. Kanneganti, *Annu. Rev. Immunol.* **38**, 567 (2020).
5. S. Fulda, D. Vucic, *Nat. Rev. Drug Discov.* **11**, 109 (2012).
6. J. Silke, P. Meier, *Cold Spring Harb. Perspect. Biol.* **5**, a008730 (2013).
7. E. C. Dueber *et al.*, *Science* **334**, 376 (2011).
8. K. Lotz, G. Pyrowolakis, S. Jentsch, *Mol. Cell. Biol.* **24**, 9339 (2004).
9. J. Ren *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 565 (2005).
10. S.-S. Liu *et al.*, bioRxiv 10.1101/2022.12.10.519866 (2022).
11. R. Jia, J. S. Bonifacino, *eLife* **8**, e50034 (2019).
12. A. Tsherniak *et al.*, *Cell* **170**, 564 (2017).
13. A. T. Nguyen *et al.*, *Science* **357**, eaan0218 (2017).
14. K. Yanagitani, S. Juszkiewicz, R. S. Hegde, *Science* **357**, 472 (2017).
15. M. Békés, D. R. Langley, C. M. Crews, *Nat. Rev. Drug Discov.* **21**, 181 (2022).

10.1126/science.adg9605

RETROSPECTIVE

Paul Berg (1926–2023)

Father of genetic engineering

By **David Baltimore**

Paul Berg, the pioneering biochemist who invented recombinant DNA technology, died on 15 February at age 96. Paul, whose work made genetic engineering possible, was a bridge between the traditional world of biochemistry and metabolism and the modern world of molecular biology.

Born in Brooklyn, New York, on 30 June 1926, Paul served in the navy during World War II and then received a bachelor's in biochemistry from Penn State University in 1948. In 1952, he earned a PhD in biochemistry from Case Western Reserve University (then Western Reserve University). He did postdoctoral work in Copenhagen in Herman Kalckar's laboratory before returning to the United States to work with Arthur Kornberg at Washington University in St. Louis. Kornberg moved his productive laboratory to the newly established medical school at Stanford University, and Paul spent the rest of his career in Palo Alto, California, in Stanford's legendary biochemistry department.

Paul's initial work was in intermediary metabolism, exploring the energetics of biochemical reactions. He made a major advance in the understanding of protein synthesis, for which he received the 1959 Eli Lilly Award in Biological Chemistry. He also examined the role of RNA in protein synthesis.

In the 1960s, Paul changed his focus from pure biochemistry to mammalian virology. He did a sabbatical at the Salk Institute in the virology laboratory of Renato Dulbecco and there learned from Marguerite Vogt how to do cell culture. His longtime research assistant, Marianne Dieckmann, accompanied him, and together they learned how to work with simian virus 40 (SV40), which contains DNA and induces cancer. In the early 1970s, as the enzymology of DNA manipulation became available, Paul recognized that the DNA of SV40 and the DNA from bacteria or bacterial viruses could be joined to form hybrids or chimeric molecules—later known as recombinant molecules. The technology of making and manipulating such non-natural but extremely valuable molecules was called recombinant DNA technology.

Such molecules were later made and inserted into bacteria by Herbert Boyer and Stanley Cohen, but the Nobel Committee for Chemistry reached back to the initiating biochemistry and recognized Paul for the discovery. Recombinant DNA methods rapidly spread throughout biology, creating a revolution in the way that questions were investigated. The techniques became the basis of the biotechnology industry.

The 1980 Nobel Prize in Chemistry recognized two techniques that revolutionized biology—half of the prize was awarded to Frederick Sanger and Walter Gilbert for sequencing DNA and the other half to Paul for recombining DNA molecules. Achieved less than 30 years after the discovery of the double-helical structure of DNA by Watson and Crick, these breakthroughs raised some alarm. When Boyer and Cohen's work was presented at a 1974 summer Gordon Research Conference, the conference attendees were so disturbed by the power and implications of the experiments that they wrote a letter, sent by the conference organizers to the US National Academy of Sciences (NAS), calling for the potential hazards of the technology to be considered.

PHOTO: JOSE MERCADO/STANFORD NEWS SERVICE

In response, Paul assembled a group of people he knew would be receptive to the concerns, including me, and the group became a committee of the NAS. Led by Paul, we called for an international meeting—because this technology was going to spread throughout the world—and asked that use of the technology be very limited until a framework of concern could be put in place.

The 1975 meeting in Monterey, California, established containment protocols under which experimentation could be initiated and included procedures for extending experiments if no hazards were detected. In the end, the hazards were largely minimal, and experimentation became commonplace, but the thoughtful development of the field gave the public confidence that the scientific community was taking seriously the need to be systematic and careful about the widespread use of a powerful new technology. Paul's deft leadership made him an international figure, the person to whom others turned when issues of concern about technology arose.

In addition to his role as a groundbreaking researcher, Paul was a great citizen of Stanford University. He was a continual advocate of the role of basic science in the education of medical students. He personally trained students and postdoctoral fellows, who remained devoted to him. At Stanford, he founded and directed the Beckman Center for Molecular and Genetic Medicine, raising much of the money for its construction himself. He lived in a house on campus and participated with joy in the life of the university.

A person of grace, elegance, and deep understanding, Paul was always available to students or the public. I spent many enjoyable hours with him engaged in conversation about science, art, society, and personal life over the years. He was devoted to his family and lost much of his zeal for life when Mildred Levy, his wife of nearly 75 years, died in 2021.

Paul had a passion for collecting art, and great works covered the walls of his house. He was a denizen of the Berggruen Gallery in San Francisco and had met many of the artists whose work he owned. Paul read widely, and we loved to share our discoveries in literature. He also loved to write. With molecular biologist Maxine Singer as his coauthor, he produced two notable books. One was a primer on genes and inheritance, and the other was an extensively researched biography of George Beadle, a Nobel laureate whose research set the stage for the molecular era. Paul was deeply concerned about political affairs and to his last days aware of the latest developments on the world stage.

With Paul's death, science has lost one of its strongest, most humane, and most accomplished supporters. The world will be harder to understand without his wisdom. ■

Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA.
Email: baltimo@caltech.edu

10.1126/science.adh2943

Algae bloom on the Baltic Sea coast near Stockholm, Sweden, in 2020.

PHOTO: PONTUS LUNDAHL/GETTY IMAGES

BOOKS *et al.*

CHEMISTRY

Phosphorus in all its forms

Limited availability and unwanted effects render the mineral's future uncertain, despite its agricultural importance

By **Robert W. Howarth**

Journalist Dan Egan's approachable new book, *The Devil's Element: Phosphorus and a World Out of Balance*, is an enjoyable, lively, and thought-provoking read targeted to the general public. Yet even I, an expert on phosphorus and the environment, learned much that I had not previously known.

Phosphorus fertilizer is critical to agriculture: The modern agricultural system that supports our global population of 8 billion people simply would not exist without it. But just as phosphorus stimulates crop growth, it can also stimulate the growth of algae and cyanobacteria, particularly in freshwaters, and excess phosphorus can lead to excessive production of these microorganisms and a global decline in water quality.

Egan's book has a strong historical perspective and is peppered with fascinating details. An early chapter presents the juxtaposition of the discovery in Hamburg, Germany, of phosphorus as an element in 1669 and the central role of phosphorus in the firebombing by Allied forces of Hamburg 274 years later in 1943.

Other chapters explore the various sources from which phosphorus has been extracted for agricultural use, starting in antiquity with manure; moving to bones, which had become a major source of the element by 1800; and later, bird excrement, or guano, which was mined on islands off of Peru in the mid-1800s. Once guano deposits became depleted, the global supply of phosphorus turned from "bones to stones," writes Egan.

Since the early 20th century, the world's phosphorus has largely come from mining high-phosphorus rock formations. The book covers some of the horrors resulting from industrial-scale mining for phosphorus, including the destruction of Banaba Island in the central Pacific Ocean, which allowed for the development of agriculture in Australia and New Zealand, and the long, bloody, and ongoing war between Morocco and the native Sahrawi people of the Western Sahara.

Beginning in the 1960s, the use of phosphorus in detergents led to rapid proliferation of algal blooms. Groundbreaking research by the late limnologist David Schindler and others clearly identified phosphorus as the culprit, leading relatively rapidly to bans on phosphorus in detergents and a remarkable improvement in water quality. However, in the decades since this success, phosphorus in lakes has again increased to destructive levels. This time the cause is agriculture. Society has so far largely failed to address the problem of nutrient pollution from agriculture.

One aspect of Egan's writing that makes his book so approachable is the large number of human-interest stories he includes. One such anecdote features a detailed description of beachcomber Gerd Simanski, who, in 2014, was badly burned by a quarter-sized "rock" collected on the shores of the Baltic Sea. The rock turned out to be pure phosphorus left over from World War II. Protected for decades in oxygen-free mud, it proved to be very combustible once placed in Simanski's pocket. He would spend most of the next 2 months recovering from the burns he sustained from the "very little rock."

Stories of David Schindler's scientific approach and achievements, as well as his early life, also enrich the book. Included is the famous story of how Schindler used a picture of his "whole-lake experiments" to show the public and policy-makers the pronounced effects phosphorus could have on freshwater ecosystems (*1*). Less well known—and previously unknown to me, although Schindler was a close friend and mentor—is Egan's charming account of Schindler's interview for a Rhodes Scholarship during his undergraduate studies. The review committee, we learn, began by focusing its questions on art history, baffling Schindler until he recognized their error: His application had listed "limnology" as his field of study, which the committee interpreted as an interest in art, on the basis of the Latin root "limn." (Limnology stems from a Greek root, not Latin, and refers to the study of freshwaters.)

The book concludes with two questions: Will world agriculture face a future crisis as we run out of phosphorus to mine? And how can we better manage phosphorus to end the crisis of overfertilized waters? These are contentious issues—and ones that deserve discussion. Egan provides a strong introduction to possible answers and encourages the reader to engage in debate.

The Devil's Element: Phosphorus and a World Out of Balance
Dan Egan
Norton, 2023. 256 pp.

In the few cases where Egan is proscriptive, I agree with his recommendations. We should end subsidies for making ethanol from corn, for example, which provides no environmental good and consumes 40% of the US corn harvest, with heavy downstream losses of both phosphorus and nitrogen.

I highly recommend *The Devil's Element*, which presents an easily digestible introduction to a major global issue. It would be a great book for a college seminar and should be read by all interested in better managing agriculture and our global environment. ■

REFERENCES AND NOTES

1. D. W. Schindler, *Science* **195**, 260 (1977).

The reviewer is at the Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA. Email: howarth@cornell.edu

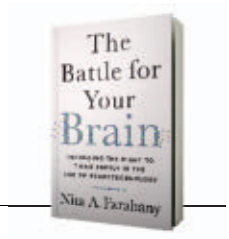
10.1126/science.adg3927

NEUROSCIENCE

Achieving cognitive liberty

Neurotechnologies necessitate new thinking on human rights

The Battle for Your Brain: Defending the Right to Think Freely in the Age of Neurotechnology
Nita A. Farahany
St. Martin's Press, 2023.
288 pp.



By **José M. Muñoz**

In her latest book, *The Battle for Your Brain*, neuroethicist and law professor Nita Farahany warns readers that neurotechnology—technology designed to monitor or manipulate the human nervous system—can "either empower or oppress us." Farahany generously illustrates how such technologies, which range from monitoring tools such as functional magnetic resonance imaging and the electroencephalogram (EEG) to techniques that can alter brain function, such as deep brain stimulation and transcranial magnetic stimulation, are affecting our lives. But *The Battle for Your Brain* is, above all, a call to expand human rights to include "cognitive liberty," a right articulated in 2004 by Wrye Sententia as our "autonomy over [our] own brain chemistry" (*1*).

Farahany expands and develops this description by proposing that in order to ensure cognitive liberty, we must update three existing human rights, the first of which is the right to privacy. Our brain data, she maintains, are highly sensitive, because they may be used to decode and infer intimate traits related to an individual's identity, emotions and feelings, intentions, memories, and even ideology. However, according to Farahany, this does not make mental privacy ("our last bastion of freedom") an absolute right. We must strike a balance, she argues, between individual and societal interests. The benefit of using an EEG device to monitor a trucker's fatigue levels to prevent traffic accidents, for example, might be worth the cost to the driver's privacy.

To establish responsible guidelines based on this balance, Farahany believes that a dialogue between academia, governments, corporations, and the public is essential. Citizens must also be guaranteed access to their brain data, and a literacy around such data must be cultivated.

The second right that Farahany maintains needs updating is the freedom of thought, which she argues is absolute—its violation being unjustified in any case. She emphasizes that this right, usually applied to the context of religious freedom, has a wide potential for expansion if we consider the capacity of neurotechnology to infer emotions and thoughts. At the same time, she cautions against adopting an excessively broad concept of freedom of thought. Attempting to discover and understand what others think is an essential function that we regularly perform, she argues; therefore, banning every kind of "mind reading" could jeopardize human coexistence.

We must prioritize our rights to autonomy, privacy, and freedom of thought as new brain monitoring and altering tools emerge.

Farahany proposes that we focus our efforts instead on preventing government use of brain wave pattern authentication—the verification of one's identity by comparing one's brain activity patterns against reference profiles in a database—as a means of social control. We have the right, she also argues, not to have our thoughts surveilled and used against us in contexts such as political opposition and criminal proceedings.

The final right that Farahany believes intersects with cognitive liberty is self-determination as it relates to our ability to access and alter our brain states. She argues that brain enhancements and diminishments—used to improve cognitive capabilities and to attenuate undesirable mental experiences, respectively—are "fundamental to human flourishing." Farahany believes that self-determination, like privacy, is not an absolute right and that societal interests may justify restricting individual brain enhancement or diminishment in certain circumstances.

Self-determination is also relative with regard to mental manipulation; persuasion is, after all, part of society. For this reason, as long as it is conducted ethically and without the intent to cause harm, Farahany is willing to concede that neuromarketing—a type of marketing that incorporates consumer brain data—does not represent a violation of cognitive liberty. (She notes, however, that this is very different from the weaponization of neuroscience and the use of psychological torture, which violate human dignity.)

Farahany ends her analysis by inviting readers to join the debate about the benefits and risks of various transhumanist proposals. These include postmortem brain cryopreservation, the expansion of human senses through brain–computer interfaces, brain-to-brain communication, brain-to-text messaging, and the use of brain implants to inactivate pain and suffering.

In *The Battle for Your Brain*, Farahany calls for "prudent vigilance and democratic deliberation" regarding the social repercussions of neurotechnology. The book is valuable reading, not only for those interested in neuroscience but also for anyone genuinely concerned about the challenges humanity will face in the near future. ■

The reviewer is at the Mind-Brain Group, Institute for Culture and Society (ICS), University of Navarra, Pamplona, Spain; the Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; and the International Center for Neuroscience and Ethics (CINET), Madrid, Spain. Email: jmunozorteg@unav.es

REFERENCES AND NOTES

1. W. Sententia, *Ann. N. Y. Acad. Sci.* **1013**, 221 (2004).

10.1126/science.adf8306

LETTERS

Edited by **Jennifer Sills**

Mitigate diseases to protect biodiversity

Action is needed to mitigate global biodiversity loss, as affirmed by the Convention on Biological Diversity. At the UN Biodiversity Conference held in December 2022 in Montreal (COP15), almost 200 nations agreed to reverse ecosystem and species loss by 2030 (*1*). During the negotiations, the risks of emerging infectious diseases were discussed in relation to humans and livestock. However, the increasing threat such diseases pose to biodiversity was overlooked.

As a result of anthropogenic activities, emerging infectious diseases spread rapidly, infecting naïve hosts that often lack effective response mechanisms, which can induce declines and extinctions (*2*). For example, Ranaviruses have caused mass mortalities in amphibians, reptiles, and fish all over the world (*3*). Fungal infections, such as white-nose syndrome in North American bats (*4*) or the recently discovered salamander plague in Europe (*5*), have led to substantial declines in their respective host populations, posing challenges to conservation efforts. These and other wildlife diseases are spread through human activities, such as travel and animal trade (*2*, *3*, *6*, *7*). New emerging infectious diseases are being discovered with increasing frequency (*2*, *5*).

Once established, it is almost impossible to mitigate emerging infectious diseases. A multilateral One Health strategy can help to eradicate them before they spread. This approach requires monitoring, early warning systems (including citizen science–based reporting), and rapid response programs across borders. Swift information transfer is crucial. Because national claims on genetic resources, stipulated by the Nagoya Protocol (*8*), could lead to a delay, exceptions should be made in the case of emerging infectious diseases.

Acknowledging the impact of emerging infectious diseases on wildlife in multinational agreements will foster global efforts to develop and execute mitigation strategies and protocols. COP15 neglected to seize this opportunity. To prevent biodiversity loss and to combat emerging infectious diseases, the Convention on Biological Diversity should work to implement the One Health approach immediately as well as incorporate the strategy into its conservation targets at the next conference.

Amadeus Plewnia, Philipp Böning, Stefan Lötters*
Department of Biogeography, Trier University, Trier, Germany.
*Corresponding author.
Email: loetters@uni-trier.de

REFERENCES AND NOTES

1. Convention on Biological Diversity, "COP15: Nations adopt four goals, 23 targets for 2030 in landmark UN biodiversity agreement" (2022); https://www.cbd.int/article/cop15-cbd-press-release-final-19dec2022.
2. P. Daszak *et al.*, *Science* **287**, 443 (2000).
3. M. J. Gray, V. G. Chinchar, Eds., *Ranaviruses: Lethal Pathogens of Ectothermic Vertebrates* (Springer, 2015).
4. D. S. Blehert *et al.*, *Science* **323**, 227 (2009).
5. A. Martel *et al.*, *Science* **346**, 630 (2014).
6. H. T. Reynolds, H. A. Barton, *Microbiol. Spectr.* **1**, 1 (2013).
7. T. T. Nguyen *et al.*, *Amphib. Reptil.* **38**, 554 (2017).
8. Secretariat of the Convention on Biological Diversity, "Nagoya Protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization" (2011); https://www.cbd.int/abs/doc/protocol/nagoya-protocol-en.pdf.

10.1126/science.aad2059

A biologist checks a little brown bat for white-nose syndrome, a disease that threatens North American biodiversity.

PHOTO: BENJAMIN OLSON/MINDEN PHOTOS

Sedimentation sifted out of pollution priorities

The Kunming-Montreal Global Biodiversity Framework (GBF) includes four goals and 23 targets to halt biodiversity loss and restore natural ecosystems by 2030 (*1*). The list includes goals to reduce pollution from sources such as plastics and nutrients (Target 7) but overlooks sediment—a key driver of poor water quality that threatens freshwater and marine ecosystems. To conserve aquatic environments, the global community must prioritize explicit indicators and commitments to reduce excess sediment.

Excess sediment is caused by land-use change and unsustainable development including logging, agriculture, and construction. When sediment enters rivers, lakes, and coastal waters, it can smother nonmobile organisms, such as plants and corals. It can also reduce the light availability and water quality necessary for many species to grow, feed, and reproduce. As a result, sediment can impede ecosystem health and function (*2*–*4*) and reduce resilience to climate change (*2*, *5*). In the hydrologic south, land use change led to a 41% increase in sediment run off between 1984 and 2021 (*6*). Globally, more than 40% of coral reefs are at risk from sediment export (*7*).

Governments and industry should work together with scientists to monitor and mitigate anthropogenic sediment impacts on freshwater and marine systems. Water quality (e.g., turbidity) and erosion metrics are relatively easy to measure through traditional and remote sensing methods and can be used to identify sedimentation (*8*). In addition to systematic land restoration and protection to combat land conversion, mitigating the negative effects of sediment requires erosion and sediment control, including maximizing covered ground; management of overland water flow; and sediment trapping, particularly in areas with high erosion risk like steep slopes.

In Australia, governments have committed to sediment reduction regulations in catchments near the Great Barrier Reef (*9*). Programs such as the UN Educational, Scientific, and Cultural Organization International Sediment Initiative (*10*) have also documented effective strategies. Similar policies should be incorporated into global pollution reduction commitments.

Managing sediment pollution would help to achieve global goals by facilitating habitat and species conservation (GBF Targets 1 to 4), sustainable food production [UN Sustainable Development Goal (SDG)

2], cleaner water (SDG 6), more responsible urbanization (SDGs 3 and 11), and better natural resource management (GBF Target 10 and SDGs 12, 14, and 15). Sediment-related policies could also increase ecosystem climate resilience (GBF Target 8, SDG 13, and the Paris Agreement) (*1*, *11*, *12*).

Caitlin D. Kuempel
Australian Rivers Institute, Griffith University, Nathan, QLD 4111, Australia.
Email: c.kuempel@griffith.edu.au

REFERENCES AND NOTES

1. Convention on Biological Diversity, "Kunming-Montreal Global Biodiversity Framework" (Convention on Biological Diversity, 2022); https://www.cbd.int/article/cop15-final-text-kunming-montreal-gbf-221222.
2. I. Donohue, J. Garcia Molinos, *Biol. Rev.* **84**, 517 (2009).
3. P. N. Owens *et al.*, *River Res. Applic.* **21**, 693 (2005).
4. K. E. Fabricius, *Mar. Pollut. Bull.* **50**, 125 (2005).
5. R. Fisher *et al.*, *Nat Commun.* **10**, 2346 (2019).
6. E. N. Dethier, C. E. Renshaw, F. J. Magilligan, *Science* **376**, 1447 (2022).
7. A. F. Suárez-Castro *et al.*, *Glob. Change Biol.* **27**, 5238 (2021).
8. J. Li, R. R. Carlson, D. E. Knapp, G. P. Asner, *Remote Sens. Ecol. Conserv.* **8**, 521 (2022).
9. Australian Government, Queensland Government, "Reef 2050 Water Quality Improvement Plan" (2018); https://www.reefplan.qld.gov.au/.
10. C. Liu *et al.*, "Sediment problems and strategies for their management" (UNESCO, 2017).
11. UN, "Transforming our world: the 2030 agenda for sustainable development" (2015).
12. UN, "Paris Agreement to the United Nations Framework Convention on Climate Change" (2015).

10.1126/science.adh2147

Outdated cap on NIH research grant budgets

In the News story "Research gets a boost in final 2023 spending agreement" (23 December 2022, p. 1263), *Science* News Staff describe the large increase in the US National Institutes of Health's (NIH's) budget for the coming year. They do not mention that, despite the increase, NIH has not raised the annual cap of $500,000 on R01-equivalent budgets in more than 25 years (*1*). NIH should increase the cap to adjust to modern-day spending power and research cost requirements.

For consumers, it takes about $900,000 today to have the same purchasing power as $500,000 in 1998 (*2*). For researchers, examination of the Biomedical Research Development Price Index (BRDPI) (*3*), the NIH's weighted metric of inflation that is driven largely by personnel costs (*4*), reveals a similar picture. Historical BRDPI trends (*3*, *5*) suggest that it took about $996,000 in 2021 to conduct research equivalent to that conducted for $500,000 in 1998. By any measure, an R01 grant cannot support nearly as much science today as it could 25 years ago. Looking ahead, the annual BRDPI is projected to be higher for 2022 to 2027 than in recent years (*3*).

Meanwhile, our institutions and NIH keep asking grantees to pay for more. At institutions, the rates of fringe benefits (which cover expenses such as retirement and health insurance) associated with grant-covered salaries continue to rise [now about $0.64 for every $1 of covered faculty salary at the University of Buffalo (*6*)], and the emerging expectation that grants (rather than the institution) will cover ever-rising tuition costs makes it harder to include graduate students in grant proposals. At NIH, the 2023 NIH Data Management and Sharing Policy (*7*) rightly addresses the importance of sharing scientific data but adds the cost of doing so to already stretched budgets (*8*). To make matters worse, NIH Institutes routinely subject funded grants to across-the-board cuts [such as the 5% cut at the National Institute on Mental Health and the 17% cut at the National Cancer Institute (*9*)].

As inflation mounts and the average annual budget of R01s edges closer to the cap (*10*), scholars may be overly optimistic [e.g., (*11*)] in estimating their budgets in ways that are detrimental to the science. It is possible to request permission to exceed the cap, but approval is far from guaranteed and requires the grantee to have their budget and key proposal elements ready months earlier than the standard submission deadline. It is time for NIH to acknowledge the marked increase in the cost of running R01-funded projects and set a higher cap.

Larry W. Hawk
Department of Psychology, University at Buffalo, Buffalo, NY, USA. Email: lhawk@buffalo.edu

REFERENCES AND NOTES

1. NIH, "Revised policy on the acceptance for review of unsolicited applications that request $500,000 or more in direct costs" (2001); https://grants.nih.gov/grants/guide/notice-files/NOT-OD-02-004.html.
2. US Bureau of Labor Statistics, "CPI inflation calculator" (2023); https://www.bls.gov/data/inflation_calculator.htm.
3. NIH Office of Budget, "Biomedical Research and Development Price Index: Fiscal Year 2021 update and projections for FY 2022–FY 2027" (2022).
4. J. Mervis, *Science* **349**, 225 (2015).
5. M. Lauer, J. Wang, D. Roychowdhury, *biorxiv* 10.1101/2022.10.07.511257 (2022).
6. University at Buffalo, "UB and research foundation rates" (2023); https://www.buffalo.edu/research/research-services/ub-rates-and-facts/ub-and-rf-rates.html.
7. NIH Office of Intramural Research, "2023 NIH data management and sharing policy" (2023); https://oir.nih.gov/sourcebook/intramural-program-oversight/intramural-data-sharing/2023-nih-data-management-sharing-policy.
8. J. Kaiser, J. Brainard, *Science* **379**, 322 (2023).
9. Rochester Institute of Technology, "NIH budget reduction table" (2021); https://www.rit.edu/pht180/nih-budget-reduction-table.
10. NIH, NIH Data Book, "Report 158: R01-equivalent grants: average size" (2022); https://report.nih.gov/nihdatabook/report/158.
11. D. M. Benjamin et al., *PLOS One* **17**, e0262862 (2022).

10.1126/science.adh0969

RESEARCH

IN *SCIENCE* JOURNALS

Edited by
Michael Funk

STRUCTURE PREDICTION

Speedy structures from single sequences

Machine learning methods for protein structure prediction have taken advantage of the evolutionary information present in multiple sequence alignments to derive accurate structural information, but predicting structure accurately from a single sequence is much more difficult. Lin *et al*. trained transformer protein language models with up to 15 billion parameters on experimental and high-quality predicted structures and found that information about atomic-level structure emerged in the model as it was scaled up. They created ESMFold, a sequence-to-structure predictor that is nearly as accurate as alignment-based methods and considerably faster. The increased speed permitted the generation of a database, the ESM Metagenomic Atlas, containing more than 600 million metagenomic proteins. —MAF

Science, ade2574, this issue p. 1123

Visualization of metagenomic structural space with predictions arranged by similarity and colored by relation to characterized proteins

KIDNEY DISEASE

GM-CSF in glomerulonephritis

Glomerulonephritis is an immune-mediated kidney disease, but the contributions of individual immune cell types is not clear. Paust *et al.* characterized pathological immune cells in samples from human patients and mice with the disease. They found that $CD4^+$ T cells producing granulocyte-macrophage colony-stimulating factor (GM-CSF) caused monocytes to promote disease by producing matrix metalloproteinase 12 and disrupting the glomerular basement membrane. Targeting GM-CSF to inhibit this axis reduced disease severity in mice, suggesting this cytokine as a potential therapeutic target for patients with glomerulonephritis. —CSM

Sci. Transl. Med. **15**, eadd6137 (2023).

NANOMATERIALS

Editing tools for layered materials

Layered metal carbides and nitrides, broadly known as MXene materials, are largely derived by the etching of the A component from $M_{n+1}AX_n$ (MAX) ternary layered compounds. Ding *et al*. developed a method to chemically alter the MAX phases through a series of topotactic transformations that enable gap opening and atom species interchange. This broadens the family of MAX materials to enable the inclusion of unconventional elements, and these in turn can be used to make additional MXene materials. —MSL

Science, add5901, this issue p. 1130

STRUCTURAL BIOLOGY

Inhibiting inhibitors of apoptosis

The ubiquitin ligase BIRC6 is an inhibitor of apoptosis (IAP). Under normal conditions, it binds to apoptotic proteases and targets these proteins for degradation, preventing cell death. This mechanism can be co-opted by cancer cells, which frequently up-regulate IAPs. Hunkeler *et al*., Dietz *et al*., and Ehrmann *et al*. present complementary structures of BIRC6 complexes that illustrate the molecular mechanisms by which this key protein mediates control of apoptosis (see the Perspective by Mace and Day). BIRC6 adopts a dimeric, horseshoe-shaped architecture with a central cavity that allows for binding to target proteases. The pro-apoptotic protein SMAC binds very tightly to the same interior site as the proteases through multiple interactions, essentially irreversibly blocking

IMAGE: LIN *ET AL.*

the ability of BIRC6 to bind substrates. The structures and supporting biochemical work in these three studies provide rich insights into the functioning of this crucial gatekeeper of apoptosis and autophagy. —MAF

Science, ade5750, ade8840, ade8873, this issue pp. 1105, 1112, 1117; see also adg9605, p. 1093

HEALTH AND MEDICINE

It's all in the tears

All cells shed extracellular vesicles (EVs), which serve as intercellular communicators by circulating in body fluids and delivering their contents to different cell targets. Their makeup reflects the cellular state of their parent cells, thus rendering them essential diagnostic and therapeutic materials. Tears are secreted by the lacrimal glands, which contain enriched biomolecules filtrated from the circulating blood. Thus, tear EVs carry information from all body organs, rendering them a rich source for disease diagnosis. Using detailed proteomic analysis, Hu *et al.* identified EV proteins deriving from 37 tissues and 79 cell types. The study constitutes an important resource for leveraging information richness in tears. —ETP

Sci. Adv. 10.1126/sciadv.adg1137(2023).

ORGANIC CHEMISTRY

Powerful approach to C–C bond formation

The aldol reaction is a powerful approach to forging carbon–carbon (C–C) bonds in both biological and laboratory organic synthesis. Rahman *et al.* report a catalytic decarboxylative aldol reaction in which small changes in the chiral ligand enable selective access to each possible isomer (four are possible) with a broad range of substrates. The demonstration of a large-scale reaction generating only carbon dioxide as the by-product highlights the practicality of the method. The resulting products were readily transformed to a range of valuable chiral building blocks. —MRG

Sci. Adv. 10.1126/sciadv.adg8776 (2023).

EVOLUTION

Warning signs

Using bright coloration to warn predators off of toxic prey, or aposematism, presents a conundrum in evolution. How do brightly colored organisms survive long enough to warn predators when they are easier to predate than their cryptic peers? Loeffler-Henry *et al*. used a large phylogeny of amphibians with known warning coloration to assess how such displays evolve. After comparing a series of models, they determined that aposematism likely appears through intermediate steps in which coloration is only visible when an organism is fleeing or intentionally displaying a hidden feature. This work demonstrates how the cost of such a trait may be circumvented through intermediary phenotypes. —CNS

Science, ade5156, this issue p. 1136

CELL BIOLOGY

Going through a phase

The gut microbiota is critical for human health. Understanding how beneficial bacteria colonize the gut enables medical interventions that promote gut health. Krypotou *et al*. discovered a mechanism that enhances the fitness of a commensal bacterium in the gut. *Bacteroides thetaiotaomicron* responded to nutrient limitation and the mammalian gut environment by sequestering a transcription factor within a membraneless compartment. This molecular condensation increased transcription factor activity and modified the transcription of hundreds of genes, including several promoting gut fitness. Thus, commensal bacteria can exploit protein condensation to colonize mammalian hosts. —SMH

Science, abn7229, this issue p. 1149

IN OTHER JOURNALS

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

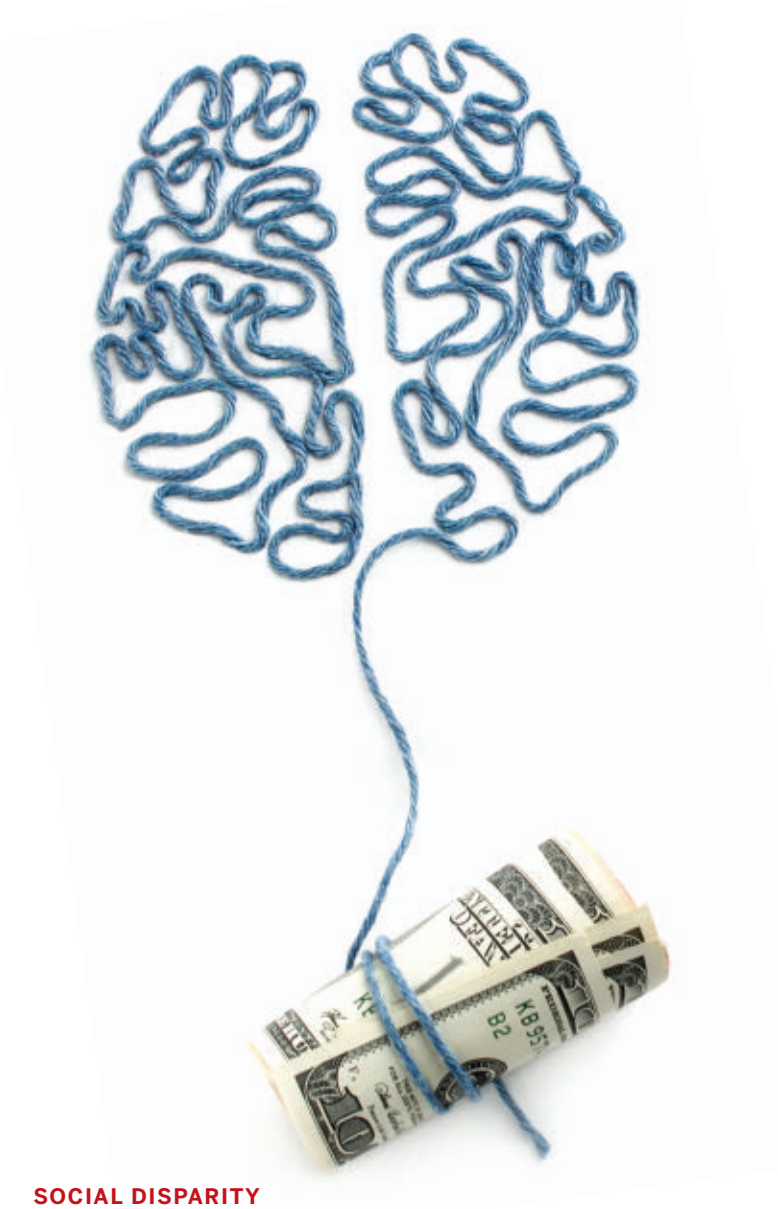


IMAGE: TANYAJOY/GETTY IMAGES

SOCIAL DISPARITY

Intelligence, wealth, and power

Should we infer unusually high intelligence in people with extremely high incomes? The highest earners make choices that are consequential because they occupy prestigious positions that wield immense power. Some may argue that the highest earners deserve their power and influence because of extraordinary intellect and merit, but this is debatable. Keuschnigg *et al.* examined cognitive ability from intellectual tests that the Swedish military required all men to take from the ages of 18 to 19. The authors then looked for a correlation between these tests and annual wages during the men's life spans from data reported to tax authorities. They found that although wages were generally higher for people with higher cognitive ability, this pattern plateaued for wages above 60,000 Euros annually. Because the highest earners were not necessarily the most intellectually gifted, other factors may have propelled them to powerful jobs. —EEU

Eur. Sociol. Rev. 10.1093/esr/jcac076 (2023).

BEHAVIOR

Timing workouts for maximum benefit

Regular physical activity has been shown to improve health and reduce the risk of cardiovascular disease and cancer. However, time of day is known to influence many metabolic parameters, so when should one exercise to optimize the benefits? Feng *et al.* investigated whether a morning or an afternoon workout has a greater health-promoting effect. Using UK Biobank data from almost 100,000 individuals, the authors showed that, as expected, physical activity was associated with reduced mortality at all ages. Mid-afternoon or mixed-time exercise, rather than morning or evening only, was associated with the lowest cardiovascular disease and all-cause mortality, particularly in aged and less fit individuals. Therefore, as with eating, there appears to be an optimal time for exercise. —MMa *Nat. Commun.* **14**, 930 (2023).

CELL BIOLOGY

Losing lysosomes by lysophagy

Selective autophagy is a process by which autophagosomes form in the cytoplasm to sequester and degrade cellular cargo. The ubiquitination of lysosomal proteins drives the selective engulfment of damaged lysosomes in a process known as lysophagy. Several adapters have been identified on the surface of damaged endolysosomes, including p62/SQSTM1. Working in HeLa cells and neurons, Gallagher and Holzbaur found that p62 functions as an essential lysophagy adapter. Loss of p62 or its ability to self-associate prevented its recruitment to damaged lysosomes and impaired engulfment by autophagosomes. Also, p62 facilitated the recruitment of the small heat shock protein HSP27, which maintained a liquid-like state of p62 oligomers, thereby promoting lysophagy. Thus, p62 facilitates lysophagy through condensate formation that is regulated by HSP27, forming a platform for autophagosome biogenesis. —SMH

Cell Rep. **42**, 112037 (2023).

GENETICS

Not-so-inactive X

Among humans and other eutherian mammals, individuals with two or more X chromosomes have one fully active X chromosome and any others are transcriptionally repressed and considered to be inactive. However, some genes on the inactive chromosome continue to be expressed, and San Roman *et al*. report that the inactive chromosome also has a direct impact on the active one. By collecting samples from patients with varying numbers of X and Y chromosomes and analyzing their gene expression patterns, the authors uncovered distinct effects of extra copies of the X, but not the Y chromosome. These effects were gene specific, which may help to explain some of the symptoms associated with X chromosome aneuploidies. —YN

Cell Genom. **3**, 100259 (2023).

BIOCHEMISTRY

Restoring NAD⁺ biosynthesis

Nicotinamide adenine dinucleotide (NAD^+) is an essential redox cofactor for metabolism, but it also serves as a cofactor for deacylation and ADP-ribosylation reactions. Declining NAD^+ levels are observed during aging, and there is a hope that restoring NAD^+ might be beneficial to health. Ratia *et al.* identified allosteric modulators of an enzyme that helps to salvage nicotinamide and determined crystal structures that reveal binding in an internal cleft near the active site. These molecules stabilized a phosphorylated form of the enzyme and reduced feedback inhibition, both functions that may help to increase the enzyme's activity. In vitro assays in cells showed an increase in NAD^+ levels for one of the compounds, a promising result for this approach. —MAF

Biochemistry **62**, 923 (2023).

INORGANIC CHEMISTRY

Electric phosphorus

Numerous commercial chemicals require the use of elemental phosphorus as a feedstock. Current methods for the reduction of naturally occurring phosphates require intensive heating to 1500°C in combination with silica and coke. Melville *et al.* explored the viability of an electrochemical reduction operating at just 800°C in a molten sodium trimetaphosphate melt. By calibrating a reference electrode to the sodium oxidation potential, the authors were able to measure overpotentials and ascertain the critical role of phosphoryl anhydrides in promoting phosphorus–oxygen bond cleavage. —JSY

ACS Cent. Sci. 10.1021/acscentsci.2c01336 (2023).

SUPERCONDUCTIVITY

Both singlet and triplet

Spin-triplet superconductors have exotic properties that make them attractive candidate materials for topological quantum computing. However, they are extremely hard to find, with only a handful of compounds having been identified so far. Recently, the material UTe_2 was found to have properties consistent with spin-triplet superconductivity. Rosuel *et al.* performed comprehensive high-magnetic-field thermodynamic measurements in UTe_2. Applying a field along a particular crystallographic axis resulted in an unusual transition between two different superconducting phases. The researchers speculate that the high-field phase is a spin-singlet and the low-field phase a spin-triplet superconductor. —JS

Phys. Rev X **13**, 011022 (2023).

PHOTO: IZUSEK/GETTY IMAGES

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CANCER

Increasing early-onset colorectal cancer

Colorectal cancer (CRC) is a common cancer type in older people, but it is increasingly being found in people below the age of 50 years, especially in people younger than 30 years. The clinical characteristics of early-onset CRC differ from those of the late-onset disease, suggesting that there is a distinct pathophysiology. In a Perspective, Giannakis and Ng discuss the differences between early- and late-onset CRC and highlight what could be driving these differences and increased incidence, which are particularly evident in high-income countries with Western diets. The authors emphasize the need for active investigation of early-onset CRC tumorigenesis and for prospective studies to identify the causal factors that could also reveal new biomarkers to improve early detection. —GKA

Science, ade7114, this issue p. 1088

OPTICAL SENSING

Enhancing optical sensing and imaging

Optical sensing and imaging can be considered as an encoding/decoding process in which the encoder is the hardware or device that takes the light signal or some property thereof (e.g., intensity, polarization, or spectral composition) and transduces that signal into usable information. The decoder is the software that then takes the information and converts it into something useful for the user. Yuan *et al*. provide a review of optical sensing and imaging methods that reflect the hardware trend toward miniaturization, reconfigurability, and multifunctional ability. Simultaneously, the development of machine learning algorithms has greatly enhanced image-processing performance. The development of both areas in concert with an information theory perspective provides a powerful platform spanning many sensing applications. —ISO

Science, ade1220, this issue p. 1103

IMMUNOLOGY

Nabbing a neoantigen

Autoimmunity can be caused by neoantigens that break immune tolerance. Zhai *et al*. profiled protein posttranslational modifications in patients with ankylosing spondylitis, an autoimmune disease (see the Perspective by Santambrogio). They found that that a cysteine residue of integrin αIIb was carboxyethylated in a process that required the gut microbe metabolite 3-hydroxypropionic acid (3-HPA) and resulted in pathogenic neoantigens. Treatment of HLA-DR4 mice with either the modified protein or 3-HPA resulted in autoantibody production and autoimmune pathology. —STS

Science, abg2482, this issue p. 1104

NEURODEGENERATION

Rescue from TDP-43 proteinopathies

Loss of the RNA-binding protein TDP-43 from the nuclei of affected neurons is a hallmark of neurodegeneration in TDP-43 proteinopathies, including amyotrophic lateral sclerosis and frontotemporal dementia. The RNA most affected by reduced TDP-43, *STMN2*, encodes stathmin-2, a protein required for axonal regeneration after injury. Baughn *et al*. found that TDP-43 sterically blocks recognition of a cryptic splice site in *STMN2* pre-mRNA (see the Perspective by O'Brien and Mizielinska). The CRISPR effector dCasRx or antisense oligonucleotides could block *STMN2* pre-mRNA cryptic splicing. This approach was able to rescue stathmin-2 levels in TDP-43–deficient human motor neurons and mouse genes edited to contain human *STMN2* cryptic splice/polyadenylation sequences. —SMH

Science, abq5622, this issue p. 1140; see also adg8501, p. 1090

CANCER IMMUNOLOGY

SUMO switches off immunosurveillance

As an extracellular cytokine, interleukin-33 (IL-33) is associated with hepatocellular carcinoma progression. However, Wang *et al*. found that as a nuclear factor, IL-33 was tumor suppressive, a function that was blocked in hepatocellular carcinoma cells. Intracellular IL-33 was SUMOylated in cell lines and tissues from hepatocellular carcinoma patients. This posttranslational modification prevented the activation of cytotoxic T cells and macrophages in vivo. Thus, antitumor immunity in hepatocellular carcinomas is partially impaired by the loss of the nuclear factor function of IL-33. —LKF

Sci. Signal. **16**, eabq3362 (2023).

CANCER IMMUNOLOGY

Overcoming a barrier in prostate cancer

Prostate cancer is minimally responsive to most immunotherapy approaches because of the poor tumor infiltration of lymphocytes. Using mouse models of prostate cancer, Zhu et al. found that cancer cell expression of the chromatin effector Pygo2 promoted immunotherapy resistance by restraining tumor T cell infiltration and cytotoxicity. Pygo2's suppressive effects were mediated by promoting the expression of the receptor tyrosine kinase Kit and the activity of indoleamine 2,3-dioxygenase 1, which occurred independently of Wnt/β-catenin signaling. Genetic deletion or pharmacological inhibition of Pygo2 enhanced prostate tumor responses to a wide range of immunotherapies. Together, these results demonstrate that Pygo2 regulates cancer cell–extrinsic immune features and represents a potential target for reducing prostate cancer resistance to immunotherapy. —CO

Sci. Immunol. **8**, eade4656 (2023).

AAAS.ORG/COMMUNITY

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

Member COMMUNITY
AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

REVIEW SUMMARY

OPTICAL SENSING

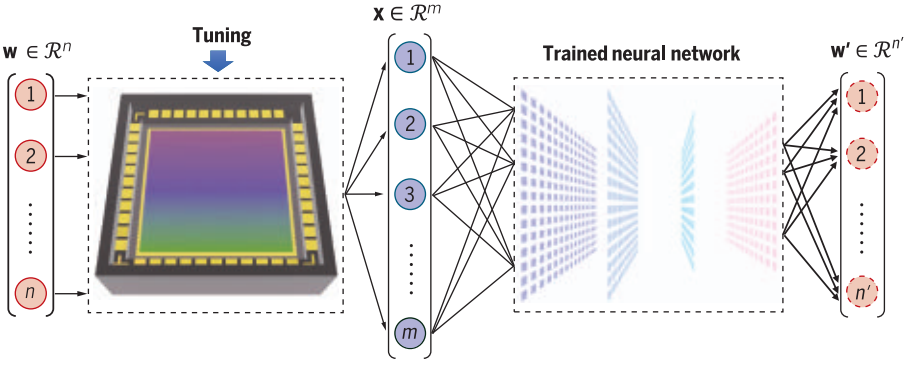
Geometric deep optical sensing

Shaofan Yuan†, Chao Ma†, Ethan Fetaya, Thomas Mueller*, Doron Naveh*, Fan Zhang*, Fengnian Xia*

BACKGROUND: Optical sensing devices measure the rich physical properties of an incident light beam, such as its power, polarization state, spectrum, and intensity distribution. Most conventional sensors, such as power meters, polarimeters, spectrometers, and cameras, are monofunctional and bulky. For example, classical Fourier-transform infrared spectrometers and polarimeters, which characterize the optical spectrum in the infrared and the polarization state of light, respectively, can occupy a considerable portion of an optical table. Over the past decade, the development of integrated sensing solutions by using miniaturized devices together with advanced machine-learning algorithms has accelerated rapidly, and optical sensing research has evolved into a highly interdisciplinary field that encompasses devices and materials engineering, condensed matter physics, and machine learning. To this end, future optical sensing technologies will benefit from innovations in device architecture, discoveries of new quantum materials, demonstrations of previously uncharacterized optical and optoelectronic phenomena, and rapid advances in the development of tailored machine-learning algorithms.

ADVANCES: Recently, a number of sensing and imaging demonstrations have emerged that differ substantially from conventional sensing schemes in the way that optical information is detected. A typical example is computational spectroscopy. In this new paradigm, a compact spectrometer first collectively captures the comprehensive spectral information of an incident light beam using multiple elements or a single element under different operational states and generates a high-dimensional photoresponse vector. An advanced algorithm then interprets the vector to achieve reconstruction of the spectrum. This scheme shifts the physical complexity of conventional grating- or interference-based spectrometers to computation. Moreover, many of the recent developments go well beyond optical spectroscopy, and we discuss them within a common framework, dubbed "geometric deep optical sensing." The term "geometric" is intended to emphasize that in this sensing scheme, the physical properties of an unknown light beam and the corresponding photoresponses can be regarded as points in two respective high-dimensional vector spaces and that the sensing process can be considered to be a mapping from one vector space to the other. The mapping can be linear, nonlinear, or highly entangled; for the latter two cases, deep artificial neural networks represent a natural choice for the encoding and/or decoding processes, from which the term "deep" is derived. In addition to this classical geometric view, the quantum geometry of Bloch electrons in Hilbert space, such as Berry curvature and quantum metrics, is essential for the determination of the polarization-dependent photoresponses in some optical sensors. In this Review, we first present a general perspective of this sensing scheme from the viewpoint of information theory, in which the photoresponse measurement and the extraction of light properties are deemed as information-encoding and -decoding processes, respectively. We then discuss demonstrations in which a reconfigurable sensor (or an array thereof), enabled by device reconfigurability and the implementation of neural networks, can detect the power, polarization state, wavelength, and spatial features of an incident light beam.

OUTLOOK: As increasingly more computing resources become available, optical sensing is becoming more computational, with device reconfigurability playing a key role. On the one hand, advanced algorithms, including deep neural networks, will enable effective decoding of high-dimensional photoresponse vectors, which reduces the physical complexity of sensors. Therefore, it will be important to integrate memory cells near or within sensors to enable efficient processing and interpretation of a large amount of photoresponse data. On the other hand, analog computation based on neural networks can be performed with an array of reconfigurable devices, which enables direct multiplexing of sensing and computing functions. We anticipate that these two directions will become the engineering frontier of future deep sensing research. On the scientific frontier, exploring quantum geometric and topological properties of new quantum materials in both linear and nonlinear light-matter interactions will enrich the information-encoding pathways for deep optical sensing. In addition, deep sensing schemes will continue to benefit from the latest developments in machine learning. Future highly compact, multifunctional, reconfigurable, and intelligent sensors and imagers will find applications in medical imaging, environmental monitoring, infrared astronomy, and many other areas of our daily lives, especially in the mobile domain and the internet of things. ■



Schematic of deep optical sensing. The n-dimensional unknown information (**w**) is encoded into an m-dimensional photoresponse vector (**x**) by a reconfigurable sensor (or an array thereof), from which **w′** is reconstructed by a trained neural network ($n' = n$ and **w′** ≈ **w**). Alternatively, **x** may be directly deciphered to capture certain properties of **w**. Here, **w**, **x**, and **w′** can be regarded as points in their respective high-dimensional vector spaces $\mathcal{R}^n$, $\mathcal{R}^m$, and $\mathcal{R}^{n'}$.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: thomas.mueller@tuwien.ac.at (T.M.); doron.naveh@biu.ac.il (D.N.); zhang@utdallas.edu (F.Z.); fengnian.xia@yale.edu (F.X.)
†These authors contributed equally to this work.
Cite this article as S. Yuan *et al.*, *Science* **379**, eade1220 (2023). DOI: 10.1126/science.ade1220

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.ade1220

REVIEW

OPTICAL SENSING

Geometric deep optical sensing

Shaofan Yuan[1]†, Chao Ma[1]†, Ethan Fetaya[2], Thomas Mueller[3]*, Doron Naveh[2]*, Fan Zhang[4,5]*, Fengnian Xia[1]*

Geometry, an ancient yet vibrant branch of mathematics, has important and far-reaching impacts on various disciplines such as art, science, and engineering. Here, we introduce an emerging concept dubbed "geometric deep optical sensing" that is based on a number of recent demonstrations in advanced optical sensing and imaging, in which a reconfigurable sensor (or an array thereof) can directly decipher the rich information of an unknown incident light beam, including its intensity, spectrum, polarization, spatial features, and possibly angular momentum. We present the physical, mathematical, and engineering foundations of this concept, with particular emphases on the roles of classical and quantum geometry and deep neural networks. Furthermore, we discuss the new opportunities that this emerging scheme can enable and the challenges associated with future developments.

Light sensors are ubiquitous and essential in many aspects of our lives. In humans, it is believed that more than 80% of the total information captured by the five senses is perceived by the eyes (*1*)—light sensors in the visible spectral range—as a result of evolution and natural selection over millions of years. There are also many different types of human-made light sensors, and such a sensor is usually built to probe a specific physical property of light. For example, an imager generates a two-dimensional (2D) map of light intensity, a spectrometer determines the spectral composition of light, and a polarimeter measures the polarization state of light. Many conventional light sensors are bulky, expensive, and monofunctional. In the past decade, as sensing tasks have become more demanding and as more computational resources have become available, two trends have emerged in optical sensing. First, it has become critical to build miniaturized, inexpensive sensors that can be integrated on-chip to enable pervasive applications, especially in mobile domains such as mobile phones, smart watches, autonomous vehicles, robots, and drones. This trend is evidenced by the demonstrations of ultracompact spectrometers in various spectral ranges (*2*, *3*) that use miniaturized dispersive optical components (*4*–*8*), on-chip interferometers (*9*–*12*), arrays of sensors with different spectral responses (*13*–*24*), or spectrally reconfigurable sensors (*25*–*28*). Moreover, on-chip polarization detectors (*29*–*31*) and compact spectral imagers (*32*–*34*) have also been extensively investigated. Second, algorithms are playing increasingly important roles in sensing, and many recent developments have leveraged machine-learning algorithms such as regression techniques and neural networks in sensor design and operation (*2*, *3*, *35*, *36*).

Here, in addition to covering some miniaturized spectral sensors, we review several innovative optical sensing schemes in which the functions of a miniaturized sensor go beyond those of traditional concepts (*37*–*47*). The recent progress in this field has been enabled by innovations in device physics and the implementation of advanced machine-learning algorithms. We approach these schemes within a common framework that we call "geometric deep optical sensing" [not to be confused with "geometric deep learning," a field that seeks to understand neural networks in non-Euclidean domains (*48*)]. The term "geometric" is intended to emphasize that the physical properties of the unknown light and the corresponding photoresponse can be regarded as points in two respective high-dimensional vector spaces and that the sensing process can be regarded as a mapping from one vector space to the other. The mapping can be linear, nonlinear, or highly entangled; for the latter two cases, deep artificial neural networks represent a natural choice for the encoding and/or decoding processes (*49*), from which the term "deep" is derived. In addition to the geometric perspective discussed above, the quantum geometry of Bloch electrons in Hilbert space, such as Berry curvature and quantum metrics, plays an important role in generating the polarization-dependent photoresponse vectors in some of the demonstrations (*47*, *50*–*60*).

[1]Department of Electrical Engineering, Yale University, New Haven, CT, USA. [2]Faculty of Engineering, Bar-Ilan University, Ramat Gan, Israel. [3]Institute of Photonics, Vienna University of Technology, Vienna, Austria. [4]Department of Physics, The University of Texas at Dallas, Richardson, TX, USA. [5]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA.
*Corresponding author. Email: thomas.mueller@tuwien.ac.at (T.M.); doron.naveh@biu.ac.il (D.N.); zhang@utdallas.edu (F.Z.); fengnian.xia@yale.edu (F.X.)
†These authors contributed equally to this work.

An information theory view

In general, from an information theory (*61*) perspective, an (optical) sensing process can be understood as follows (Fig. 1A): A sensor acts as an encoder that converts unknown, high-dimensional physical quantities into sensor outputs; a channel corresponding to a noisy measurement process reads the sensor outputs; and a decoder deciphers the encoded high-dimensional information. Here, the high-dimensional physical quantities can be characterized by a vector $\mathbf{w}$, which represents the intrinsic physical properties of a light beam, such as power, spectrum, polarization state, spatial or temporal properties, or the combination of several of these. The vector $\mathbf{w}$ can be treated as a point in a vector space of dimension n ($\mathbf{w} \in \mathcal{R}^n$) (Fig. 1B). In traditional sensing schemes, direct determination of such a vector requires a series of measurements that use different types of optical components such as beam splitters, waveplates, filters, dispersive gratings, and power meters, followed by data processing steps. In the sensing scheme introduced here, $\mathbf{w}$ is first encoded into a response vector $\mathbf{x}$ by a single sensor or an array thereof, which is engineered to capture spatial, spectral, polarization, and/or temporal information. Vector $\mathbf{x}$ can be treated as a point in a vector space of dimension m ($\mathbf{x} \in \mathcal{R}^m$) (Fig. 1B). It may be interpreted directly to capture certain properties of $\mathbf{w}$ or decoded into a vector $\mathbf{w}'$ to reconstruct the desired physical quantities of $\mathbf{w}$.

In contrast to traditional sensing schemes, in the geometric deep optical sensing scheme, both the encoding and decoding processes can be implicit. Moreover, different kinds of information can be encoded concurrently into the sensor outputs. As a result, this sensing scheme allows for the detection of multiple physical properties of light and functionality multiplexing. Depending on the dimensionality of the two vector spaces $\mathcal{R}^n$ and $\mathcal{R}^m$, we distinguish between three cases. The case $m = n$ is the most common and corresponds, for example, to the case of computational spectrometers (*13*, *18*, *20*, *25*). The case $m < n$ corresponds to compressed sensing (*62*–*65*). By using prior knowledge or proper assumptions, high-dimensional information can be reconstructed from a low-dimensional photoresponse. Finally, the case $m > n$ may have the advantage of being more robust to noise because of redundancy introduced in the information-encoding process (*61*).

Information-encoding mechanisms

As shown in Fig. 1B, the encoder's role is to map the physical information in $\mathcal{R}^n$ to a photoresponse in $\mathcal{R}^m$. In the following sections, we

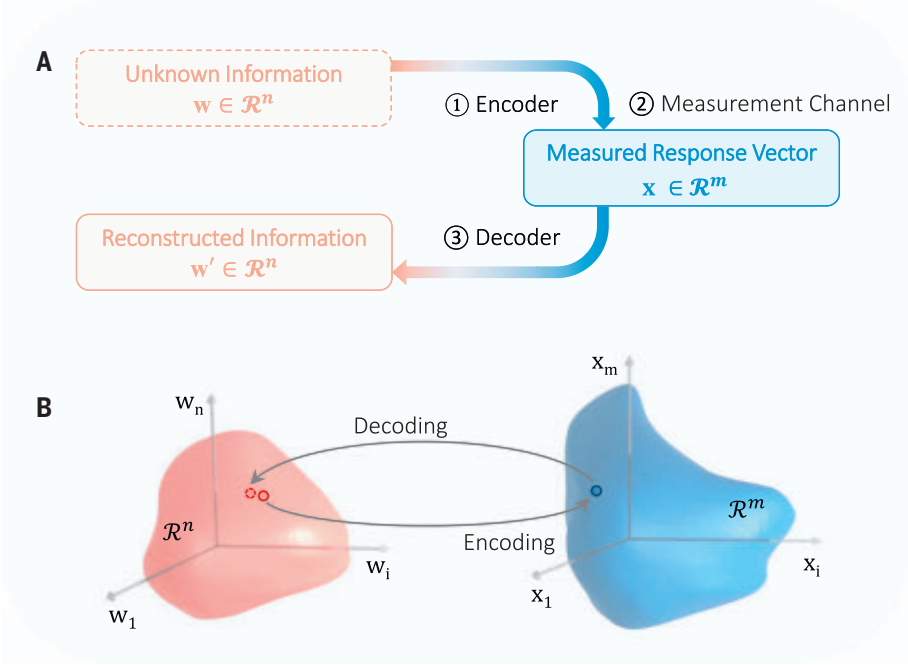


Fig. 1. An information theory view of deep optical sensing. (**A**) An encoder (an optical sensor or sensor array) converts the unknown *n*-dimensional physical information **w** ∈ $\mathcal{R}^n$ into electrical outputs, the channel corresponds to a noisy measurement process that reads the *m*-dimensional output **x** ∈ $\mathcal{R}^m$, and a decoder reconstructs the information **w′** ∈ $\mathcal{R}^n$. (**B**) The vectors **w** and **x** can be regarded as points in *n*- and *m*-dimensional vector spaces $\mathcal{R}^n$ and $\mathcal{R}^m$, respectively. A mathematical tool, for example, a trained neural network, maps **x** to **w′** from $\mathcal{R}^m$ back to $\mathcal{R}^n$. A high-performance sensor captures the unknown information accurately such that the reconstructed **w′** is close to **w** in $\mathcal{R}^n$. **w** and **w′** are represented by the red hollow dots with solid and dotted edges, respectively, in $\mathcal{R}^n$. **x** is represented by the blue hollow dot in $\mathcal{R}^m$. Alternatively, **x** can be evaluated directly to capture certain features of **w**.

discuss examples of how this can be accomplished for different properties of incident light.

Tuning device geometric features and quantum geometry for polarization encoding

Geometry plays a critical role in light-matter interactions. For example, optical devices with different geometric features can exhibit distinct polarization-dependent responses to light. Previously, polarization imaging has been demonstrated by using a 2D grating matrix (*66*) consisting of optical elements with different geometries (Fig. 2A), in which different polarization components of an incident light beam are separated spatially for the subsequent polarization information–encoding process.

Other than the geometric features of the sensor, the quantum geometry of Bloch electrons, that is, Berry curvature and quantum metric, can also be tuned for the encoding of polarization information. Quantum geometry represents the geometry of the quantum states in Hilbert space, and it is critical for nonlinear photoresponses such as the second-harmonic generation and the bulk photovoltaic effect (BPVE). One notable feature of the BPVE is its strong polarization and wavelength dependencies. In the device reported by Ma *et al.* (*47*), the active material is twisted double-bilayer graphene (TDBG) sandwiched between two hexagonal boron nitride (hBN) thin films. The graphene (top) and silicon (bottom) gate electrodes are used to produce two electric potentials through hBN to modulate the quantum geometric properties of TDBG for polarization encoding.

Under excitation with linearly polarized light, the BPVE (shift current) is determined by two independent conductivity elements σ_{xxx} and σ_{yyy} that can be directly calculated by integrating $S_{xxx(yyy)}$, which is the contribution from an electron-hole pair that participates in the resonant optical transition (*51*–*54*). Here, $S_{xxx(yyy)}$ depends on the Fermi distribution difference between the electron and hole Bloch bands, the interband non-Abelian Berry connections, and the excitation light frequency. The integrand S_{xxx} in the moiré Brillouin zone is shown in Fig. 2B for Fermi energy $E_F = 0$ and interlayer potential difference $\Delta V = 100$ meV under excitation with 7.7-μm light in TDBG (*47*). Hotspots in Fig. 2B indicate the positions in momentum space at which the quantum geometric properties relevant to σ_{xxx} are pronounced, which allows for the resonant optical transition. Under circularly polarized light excitation, the BPVE (injection current) is governed by the interband Berry curvature dipoles or the Hermitian metrics (*54*).

Importantly, not only does the nonlinear response depend on the polarization of the light, but this dependence is also tunable. The tunability originates from the fact that the two gate voltages can independently control the Fermi energy (or carrier density) and the out-of-plane displacement field of the 2D material in the device. The displacement field can modulate the band structure that specifically determines the quantum geometric properties of Bloch states. The Fermi energy determines the electron-hole pairs of Bloch states that are available for the resonant optical transition. Together, they determine the nonlinear conductivity tensor and hence tune the BPVE. As a result, the polarization information of the incident light can be encoded into the nonlinear photoresponse map that is generated under different pairs of biasing gate voltages (*47*). Such an encoding process is implicit because of the complexity of the device, including strain, disorder, inhomogeneity, and so on (*67*). However, the decoding can still be successfully performed by using a trained neural network, as discussed below.

Engineering the spectral response for optical spectroscopy

Optical elements with different geometric features can also be directly integrated with complementary metal-oxide semiconductor (CMOS) sensors to encode the spectral information (Fig. 2C) (*19*). In such a spectrometer, each element captures certain spectral characteristics by leveraging the rationally designed geometric features. Indeed, this approach has been extensively used for information encoding in spectral and polarization imaging (*17*, *19*, *68*–*70*). However, despite the effectiveness of such a geometric approach, the physical layout of optical elements can hardly be reconfigured after fabrication, which limits their potential in advanced applications. For example, to achieve high resolution in spectroscopy, a large number of elements with different geometric features are required, yet their scaling is limited by optical diffraction, which results in a large overall device footprint. Moreover, because of the lack of reconfigurability, it is difficult to fully leverage the capacities of machine-learning algorithms for deciphering nontrivial high-dimensional data.

In addition to using elements with different geometric features, there are a number of other approaches for encoding spectral information for optical spectroscopy. A prime example is the engineering of the bandgap of semiconductors; bandgap determines the photon energies at which transitions between bands can occur. Consequently, the tuning of the bandgap can enable the encoding of spectral information into a photoresponse vector. Miniaturized spectrometers have been demonstrated based on bandgap tuning by varying

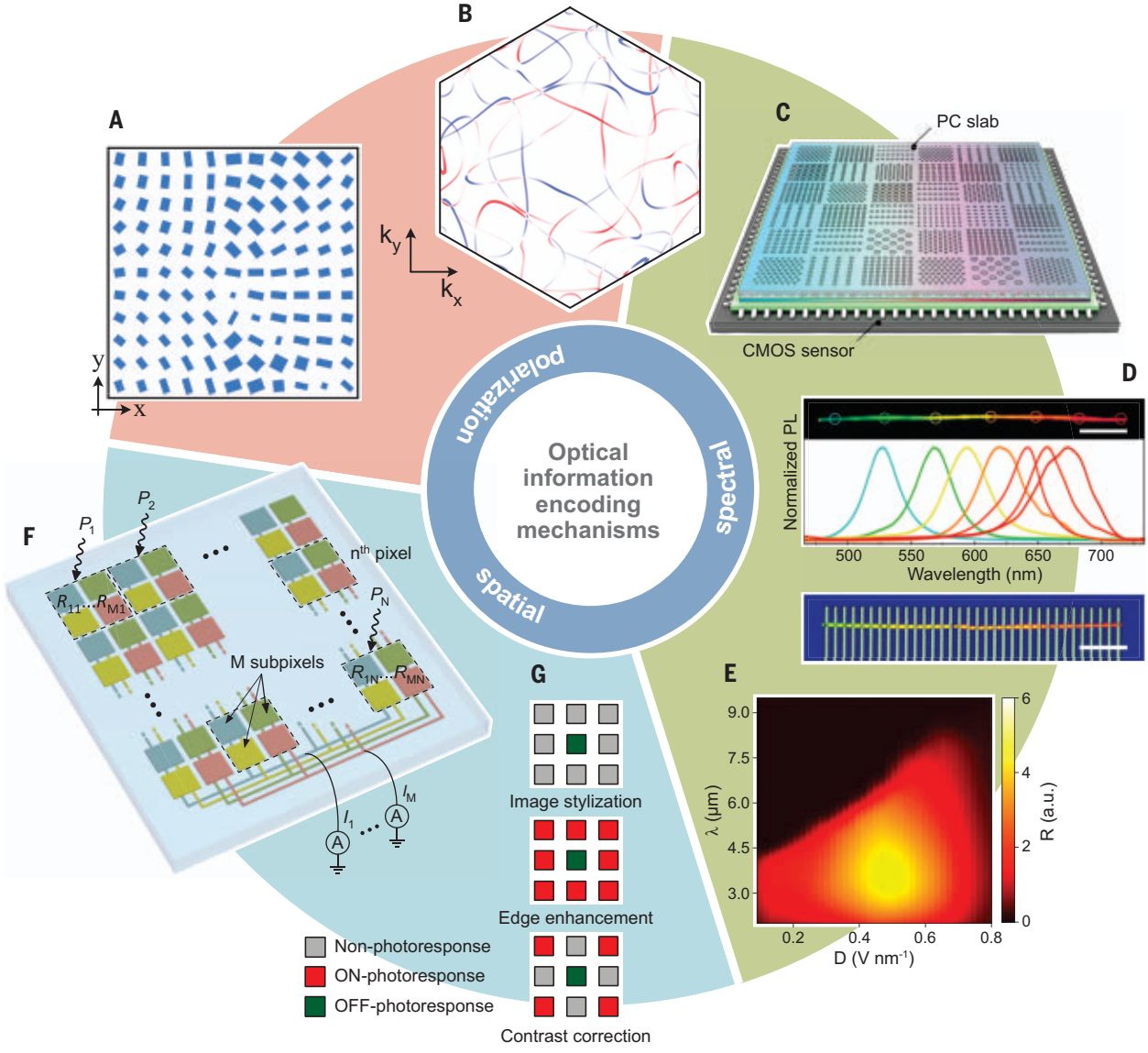


Fig. 2. Information encoding mechanisms. (**A**) Encoding the polarization information of light using a grating matrix with different physical geometries. [Adapted with permission from (*66*)] (**B**) Calculated distribution of the integrand for computing shift current of TDBG in the moiré Brillouin zone. Its tunability enables the encoding of the polarization, wavelength, and power information of mid-infrared light using the moiré quantum geometry of TDBG. [Adapted by permission from Springer Nature Customer Service Center GmbH, Springer *Nature* (*47*), copyright (2022)] (**C**) Encoding the spectral information of light using an array of photonic crystals (PCs) with different geometric features. [Adapted by permission from Springer Nature Customer Service Center GmbH, Springer *Nature* (*19*), copyright 2019)] (**D**) Bandgap tuning by varying the chemical composition for encoding spectral information. The photoluminescence (PL) spectra taken in different locations of a CdS_xSe_{1-x} nanowire with a graded composition are shown at the top. The shift in peak wavelength indicates the varying bandgap along the wire. A fabricated single-nanowire spectrometer is shown at the bottom. [Adapted with permission from (*20*)] (**E**) Spectral responsivity of a reconfigurable black phosphorus sensor under different biasing displacement fields. [Adapted by permission from Springer Nature Customer Service Center GmbH, Springer *Nature* (*25*), copyright (2021)] (**F**) Photodetector array for encoding optical images with *n* pixels into *m* electrical outputs. Each subpixel is reconfigurable by two split gates, which are biased with voltages of opposite polarities. [Adapted by permission from Springer Nature Customer Service Center GmbH, Springer *Nature* (*37*), copyright (2020)] (**G**) Photoresponsivity distributions of a reconfigurable pixel array for simultaneous image capture and processing (image stylization, edge enhancement, and contrast reduction). [Adapted with permission from (*39*)]

the chemical compositions of materials or the electrical displacement fields (*13*, *20*, *25*). In one case, different chemical compositions were introduced within a single cadmium sulfide selenide (CdS_xSe_{1-x}) nanowire to encode the spectral information of the incident light (*20*). The top panel of Fig. 2D shows the photoluminescence spectra taken at different locations along such a single nanowire, and the bottom panel illustrates a fabricated single-nanowire spectrometer. The spectral information of the light is encoded into the photoresponse vector that is measured along the wire. Moreover, the chemical composition and the dimension of quantum dots can be tuned together to cover a broad spectral range, as demonstrated in a quantum dot spectrometer (*13*).

An electric field can tune the absorption edge of a bulk semiconductor, which is well known as the Franz-Keldysh effect (*71*). In semiconductor quantum wells, such absorption

tuning is more pronounced and is referred to as the "quantum confinement Stark effect" (*72*–*75*). Recently, a single on-chip black phosphorus device has been shown to function as a mid-infrared spectrometer within the 2- to 9-μm wavelength range (*25*). The bandgap of ~10-nm-thick black phosphorus was tuned by an external electric field, by using the Stark effect for the encoding of spectral information. Figure 2E shows the spectral responsivity of a reconfigurable black phosphorus sensor (*25*) in which the cutoff wavelength has been extended to around 9 μm under a biasing displacement field of 0.8 V/nm. This demonstration has inspired further demonstrations of spectrometers based on single, reconfigurable photodetectors (*26*, *27*). In these works, the electric field does not appreciably tune the bandgap but rather changes the spectral response of the devices by adjusting the relative band alignment (*26*) or ion migration properties (*27*). Finally, metasurfaces on graphene exhibit strong tunability in the mid- and far-infrared wavelength regimes (*76*, *77*). Spectroscopy has also been demonstrated by combining such reconfigurable metasurfaces with discrete infrared photodetectors (*78*). In this case, the encoding process is realized by tuning the reflection spectra of the metasurfaces with an external electric field.

Encoding optical images with reconfigurable detector arrays

In addition to spectral and polarization information, the spatial variations of light intensity (optical images) can also be encoded in the photoresponse. To this end, an array of reconfigurable photodetectors as an image sensor and an artificial neural network have been used for ultrafast machine vision (*37*). The image-encoding process captures the spatial features directly and reduces the transmission bandwidth requirements. Figure 2F shows an illustration of the device, which consists of n photoactive pixels arranged in a 2D array, with each pixel divided into m subpixels (*37*). Each subpixel is composed of a WSe_2 photodiode whose responsivity can be reconfigured by two split gates. Enabled by the reconfigurability on the subpixel level, the sensor can be trained to encode optically projected images into an m-dimensional output, as will be further discussed in the section "Roles of neural networks in optical sensing."

The concept of capturing spatial features directly in the light-detection process has been used in several other works (*38*–*46*). For example, Fig. 2G illustrates three configurations of the responsivity matrix of a 3-pixel–by–3-pixel sensor array for image stylization, edge detection, and contrast correction in (*39*). Also in this case, different geometric features can be captured directly in the imaging process with different configurations of the device response, thus eliminating the need for subsequent computational image processing steps. In another example (*79*), a large fraction of the sensor elements in an imaging device were physically combined into several "superpixels" that extended over the entire surface area of the chip. For a given pattern recognition task, their optimal shapes were determined by using a machine-learning algorithm from training data. Classification of optically projected images on an ultrafast time scale and with an enhanced dynamic range was demonstrated.

General considerations for information encoding

We have shown that there are a number of pathways to realize information encoding. This naturally leads to the following question: What are the essential requirements for constructing a good encoder and a measurement channel? To reconstruct the physical information in $\mathcal{R}^n$ or to capture the features of interest directly from the m-dimensional photoresponse, degeneracy is not desirable. When multiple points in $\mathcal{R}^n$ are mapped to the same point in $\mathcal{R}^m$, loss-less reconstruction is no longer possible. Indeed, it is the degeneracy in their photoresponse that hinders conventional nonreconfigurable detectors from sensing richer information of unknown light because light beams with different combinations of physical properties (i.e., power, spectrum, polarization, angular momentum, geometric features, and so on) can yield the same output signal. The reconfigurability can eliminate degeneracy by increasing the dimensionality of the response. By configuring the geometric features, spectral response, and quantum geometric properties of the sensor, the tunability in its optical response can map different points in $\mathcal{R}^n$ to distinct points in $\mathcal{R}^m$. Noise introduced in the measurement process may, however, increase the possibility of overlapping originally distinctive points in $\mathcal{R}^m$, which leads to potential degeneracy. Therefore, a sufficiently large signal-to-noise ratio is important in the encoding process. This observation is analogous to the channel capacity in coding theory, where capacity increases as noise decreases. Other than noise reduction, introducing redundancy in the measurements ($m > n$) may mitigate the degeneracy problem (*61*).

Decoding pathways

Although the photoresponse vector $\mathbf{x}$ itself may contain valuable information about certain features of $\mathbf{w}$, a decoder is generally needed to decipher and reconstruct the original physical information to complete the sensing process. In this section, we discuss two general classes of models that are used in optical sensing to map the sensor response in $\mathcal{R}^m$ to the original physical information in $\mathcal{R}^n$ (or to capture some information of interest directly from $\mathbf{x}$): analytical (Fig. 3A) and data-driven (Fig. 3B). Analytical models require a comprehensive understanding of the encoder, whereas data-driven models usually use neural networks with experimental photoresponse data. In addition, other approaches exist that do not belong to either of the above two models but could represent alternative future pathways for decoding optical information. For example, randomly initialized neural networks without training have been shown to be effective in image generation and restoration (*80*, *81*), and an analytical algorithm followed by a convolutional neural network has been used to solve inverse problems (*82*).

For an exemplary illustration of an analytical decoding process, let us consider the following method, which is widely used in spectral sensing. In the linear response regime under optical excitation with spectrum P_λ, the response of a photodetector can be written as $I_S = \int R_\lambda P_\lambda \mathrm{d}\lambda$, where R_λ is the spectral responsivity (*13*–*15*, *18*, *20*, *21*). The light spectrum and the spectral responsivity can be represented in a vector space $\mathcal{R}^n$ by two vectors $\mathbf{p}_\lambda$ and $\mathbf{r}_\lambda$, respectively. The photoresponse I_S is then the inner product of these two vectors, $I_S = \mathbf{r}_\lambda^{\mathrm{T}}\mathbf{p}_\lambda$. [Similar equations hold for other linear optical properties, for example, the spatial intensity variation, $I_S = \mathbf{r}_x^{\mathrm{T}}\mathbf{p}_x$, where $\mathbf{p}_x$ represents a flattened optical image and $\mathbf{r}_x$ is a spatially varying photoresponsivity (*37*, *79*).]

Consider an optical sensor consisting of $m = n$ states, S_i, where $i = 1, 2, \ldots, n$. These n states, or measurements, may be realized by using n different subelements in the sensor or by n different operational modes of a single reconfigurable sensor. In either case, they can be represented as n inner products discussed above or as a matrix-vector product with photoresponsivity matrix $\mathbf{R}$, $\mathbf{i}_S = \mathbf{R}\mathbf{p}_\lambda$. Here, $\mathbf{R}$ is an n-by-n matrix, and its element at ith row and jth column, $R_{S_{i,j}}$, represents the discrete photoresponsivity at wavelength λ_j in state S_i; $\mathbf{i}_S$ denotes the discretized photoresponse vector $(I_{S_1}, I_{S_2}, \ldots I_{S_n})^T$; and the spectrum is denoted as $\mathbf{p}_\lambda = (P_{\lambda_1}, P_{\lambda_2}, \ldots P_{\lambda_n})^T$. If the response matrix $\mathbf{R}$ is known, then the spectrum $\mathbf{p}_\lambda$ can be reconstructed using the measured photoresponse $\mathbf{i}_S$. However, direct reconstruction by calculating the inverse $\mathbf{R}^{-1}$ may lead to unsatisfactory results because $\mathbf{R}$ may be ill-conditioned and $\mathbf{i}_S$ may exhibit measurement noise or even errors. In recent demonstrations, this problem has been mitigated using adaptive regression methods with Tikhonov (*83*) or LASSO (least absolute shrinkage and selection operator) regularizations (*84*) by minimizing a cost function, cost = $||\mathbf{R}\mathbf{p}_\lambda - \mathbf{i}_S||^2 + \alpha w(\mathbf{p}_\lambda)$. Here, α is a parameter that controls the regularization strength and $w(\mathbf{p}_\lambda)$ is a penalty function. These regularization approaches allow us to alleviate the negative effects of an ill-conditioned matrix $\mathbf{R}$ and measurement noise. In the general case, $n \neq m$, adaptive regression methods can still

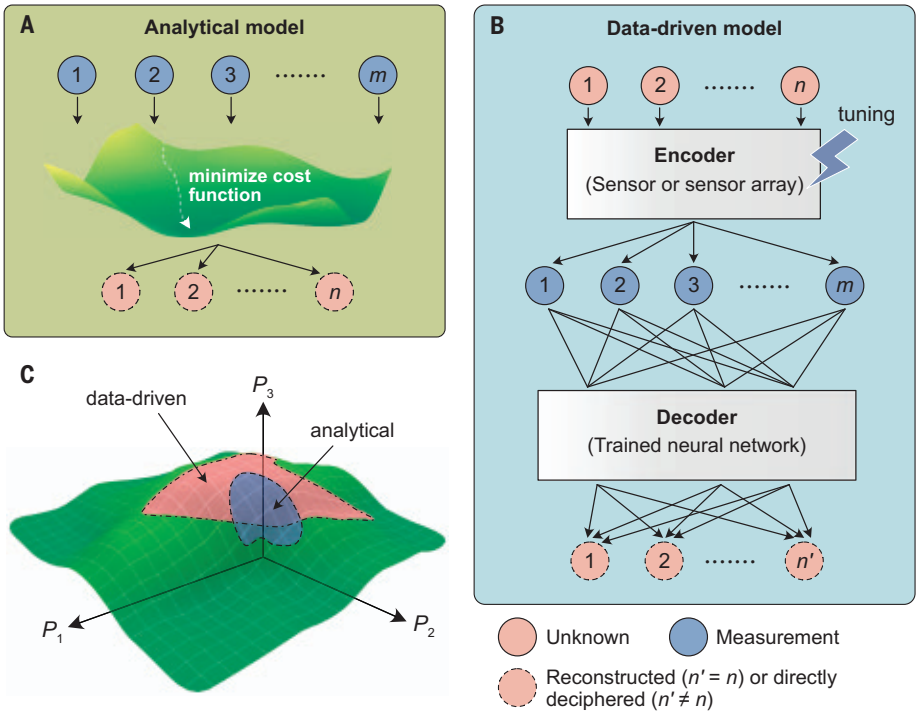


Fig. 3. Information decoding pathways. (**A**) Schematic of using an analytical approach to extract *n*-dimensional information from an *m*-dimensional photoresponse vector if the encoding process (not shown for simplicity) can be modeled explicitly. (**B**) Schematic of using a data-driven model in sensing. A reconfigurable device (or an array thereof) is used as an encoder to generate an *m*-dimensional photoresponse vector, and a trained neural network is used as a decoder to decipher the *n*-dimensional information ($n' = n$). The sensor itself can also act as (part of) a neural network. Moreover, desirable features of the *n*-dimensional information may be directly extracted from the *m*-dimensional photoresponse. (**C**) Schematic of a manifold in a conceptual, 3D parameter space. For analytical models, usually a smaller parameter space (blue surface) is required to capture the information, whereas a larger parameter space (red surface) is required for data-driven models.

be applied to compute a solution that minimizes the cost function.

Unlike analytical models, data-driven sensing generally consists of two steps. First, the model, usually a neural network, needs to be trained. This may occur on-chip or off-chip, in a supervised or an unsupervised or self-supervised manner (*37*–*42*, *47*). In the former case, both a set of sensor inputs $\mathbf{w}$ and their photoresponse vectors $\mathbf{x}$ (or directly deciphered information, e.g., classification results) are provided. In the latter, an efficient representation (encoding) is learned from a set of inputs $\mathbf{w}$ alone, and the decoder attempts to reproduce at its output the original information, $\mathbf{w}' \approx \mathbf{w}$. After training, the neural network can then be leveraged to decipher the unknown information based on the measured photoresponse vector $\mathbf{x}$.

Data-driven models have several distinctive advantages that make them suitable for advanced sensing applications. First, data-driven models can be used as decoders to exploit existing experimental results, even when analytical models are not accessible. As illustrated in Fig. 3B, after the neural network is trained, the response can be interpreted without involving any specific mathematical relation. Second, different types of physical information can be deciphered simultaneously, as long as the training process takes them into account. Third, it is possible to realize functionality multiplexing because the outputs of the data-driven models are not limited to specific physical properties of light. For example, imaging and classification functions can be combined using trained neutral networks, which substantially reduces the complexity of the overall system.

At the same time, data-driven models require the acquisition of sufficient training data, which need to be correctly labeled in the case of supervised learning. If their characteristics do not change substantially during operation, then the sensors only need to be trained or calibrated once by the manufacturers without the end users having to go through this process. Data augmentation methods, such as interpolation and data synthesis, can be used to expand training datasets (*85*). In addition to the initial training of the model, recalibration during operation can also be applicable to sensors in both analytical and data-driven models. Choosing reliable references is crucial for deploying sensors in recalibration because laboratory-level calibration may not be available. Good references should have specific and stable features, and measuring them will provide enough information for recalibration. Examples include checkerboard and US Air Force (USAF) 1951 targets for imaging as well as elemental and molecular spectral lines for spectroscopy. Recalibration should focus on parameters directly related to drifting and degradation, which require a comprehensive understanding of the sensor's physical properties.

Manufacturing and environmental variations and measurement noise are further issues that need to be considered. Sensors may be sensitive to manufacturing variability and environmental conditions, such as temperature, humidity, and stray light. Advanced packaging schemes can increase tolerance to these conditions. The effects of these variations can be compensated numerically, if well understood. Measurement noise can be minimized by optimizing encoding processes and taking into account the sensor's physical properties and sensing requirements. Designing application-specific encoding strategies is critical to achieve both efficiency and accuracy. For example, focusing on operational states that are strongly affected by targeted spectral features can improve the performance of spectral sensing.

Regardless of the model used, the dimensionality m of the photoresponse vector is determined by the measurement parameter space. The parameters used to characterize the photoresponse vector can be diverse. For example, in tunable dual-gate sensors (*25*, *47*), top (V_{TG}) and bottom (V_{BG}) gate biases are typical parameters that together form a 2D parameter space. Bias voltage in the photocurrent generation path, ambient temperature, load applied to the sensor, and external magnetic field may also be among the parameters for photocurrent measurements, depending on the sensing application. A conceptual, 3D parameter space is shown in Fig. 3C, as illustrated by axes P_1, P_2, and P_3. In practice, the parameter space can be reduced by a set of constraints to a manifold (green surface in Fig. 3C) on which the photoresponse is measured. When analytical models are applied, the understanding on the sensor is usually extensive. As a result, a smaller parameter space can be used. For example, in a previously demonstrated dual-gate black phosphorus spectrometer, it is known that at charge-neutrality, the photoresponse is highest and the bandgap tuning is effective (*25*). As a result, it is not necessary to perform the photoresponse measurements across the entire 2D parameter space of V_{TG} and V_{BG}, but they can instead be performed in a 1D line along which V_{TG} and V_{BG} collectively induce no net doping. In a data-driven model, the understanding of the sensor does not necessarily need to be as comprehensive. In this case, usually a larger parameter space is needed

to fully capture the information of unknown light, as illustrated in (*47*), in which a 2D parameter space of V_{TG} and V_{BG} was used. The larger (red) and smaller (blue) areas in Fig. 3C schematically represent these two types of parameter spaces, respectively.

Roles of neural networks in optical sensing

We now discuss potential roles of artificial neural networks in the information encoding and decoding in optical sensors. Autoencoders are a specific class of neural networks that are particularly promising for the sensing scheme introduced here. Figure 4A shows a schematic of the network (*37*). An autoencoder consists of two parts: an encoder that compresses the input data in a bottleneck layer with $m < n$ dimensions and a decoder that attempts to reproduce the original data at its output. Figure 2F shows an illustration of a device realization for image encoding (*37*). We emphasize, however, that the concept is not limited to imaging alone; it can also be applied to other optical sensing tasks, such as spectral measurements. The device consists of $n = 9$ photoactive pixels arranged in a 2D array, and each pixel is divided into $m = 3$ subpixels. Each subpixel consists of a WSe_2 photodiode whose photoresponsivity can be configured to adjust the synaptic weights. By interconnecting the subpixels, an integrated neural network can be formed in which the encoder is the optical sensor itself and the decoder is the external computer programs. Figure 4B shows the operation of a device after a training process that is based on backpropagation (*37*). The encoder translates the projected images (letters "n," "v," and "z") into an output current vector, which is converted by a nonlinearity into an activation code and finally reconstructed into the original image by the decoder. It is noticeable in Fig. 4B that the activation codes are binary. This is a consequence of the training process, in which Gaussian noise was injected to learn binary representations. Such a device may be operated as a binary-hashing autoencoder, which eliminates the need for analog-to-digital conversion before signal reconstruction. To implement a deep autoencoder, additional hidden layers can be added to both the encoder and the decoder to deepen the network and increase its complexity and performance. Although this is straightforward on the (software) decoder side, on the (hardware) encoder side, it is more elaborate but could be achieved, for example, by converting the output currents to voltages that are then fed into a memristor crossbar (*86*, *87*). However, as demonstrated in Fig. 4C, comparable performance can be achieved by keeping a single layer for the encoder and only increasing the number of layers of the decoder.

Other neural network architectures can also be implemented using a similar device structure. A machine vision processor was developed to operate as a convolutional neural network by integrating 1024 MoS_2 photo–field effect transistors in a crossbar structure, and a classification of digits from the Modified National Institute of Standards and Technology (MNIST) dataset was demonstrated (*38*). Figure 4D illustrates the working principles of another classifier presented in (*39*) by implementing a convolutional neural network with a prototypical 3-pixel–by–3-pixel sensor. Here, the photoresponsivity of each pixel is reconfigurable, and the total photocurrent represents the convolution of the image and the responsivity matrix. Binary figures representing the letters "n," "j," and "u" were used during the training to obtain the responsivity matrix of each letter. A testing classification accuracy of 100% was achieved by using the weighted average of the convolutional kernel. In a third example (*42*), an array of black phosphorus programmable phototransistors that can be programmed with 5-bit precision was used to implement an in-sensor convolutional neural network.

In addition to the image recognition and processing functions described above, it is also possible to detect multiple physical properties

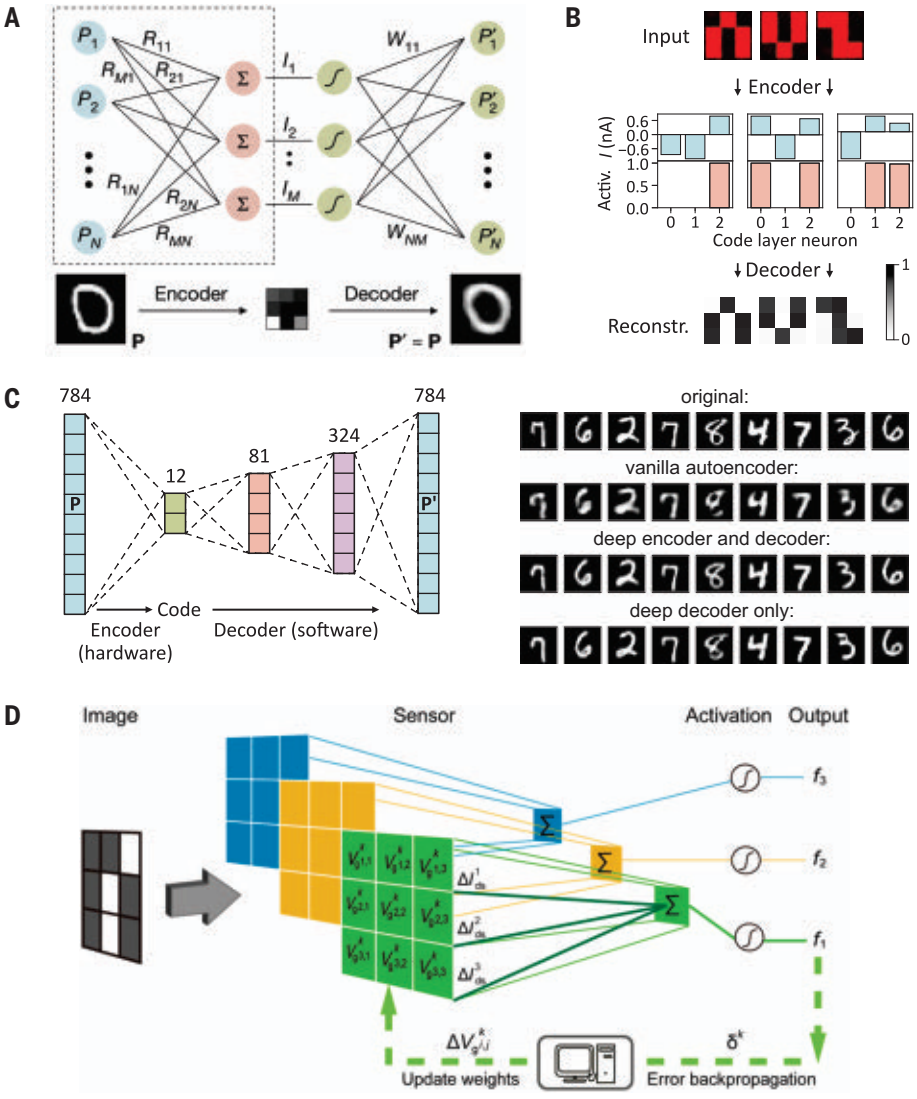


Fig. 4. Machine vision using a reconfigurable sensor array. (**A**) Illustration of an autoencoder with a single hidden layer. The bottom shows the encoding and decoding of a letter from the MNIST database. (**B**) Operation of an autoencoder based on a reconfigurable 9-pixel WSe_2 sensor array. The sensor array acts as an encoder that translates images into current codes that can later be reconstructed into the original image by an external decoder. [(A) and (B) are adapted by permission from Springer Nature Customer Service Center GmbH, Springer *Nature* (*37*), copyright (2020)] (**C**) Illustration of an autoencoder with a deep decoder (left) and MNIST image ($n = 784$) reconstruction ($m = 12$) using different types of autoencoders (right). (**D**) Convolutional neural network realized by a reconfigurable 9-pixel retinomorphic vision sensor. [Adapted with permission from (*39*)]

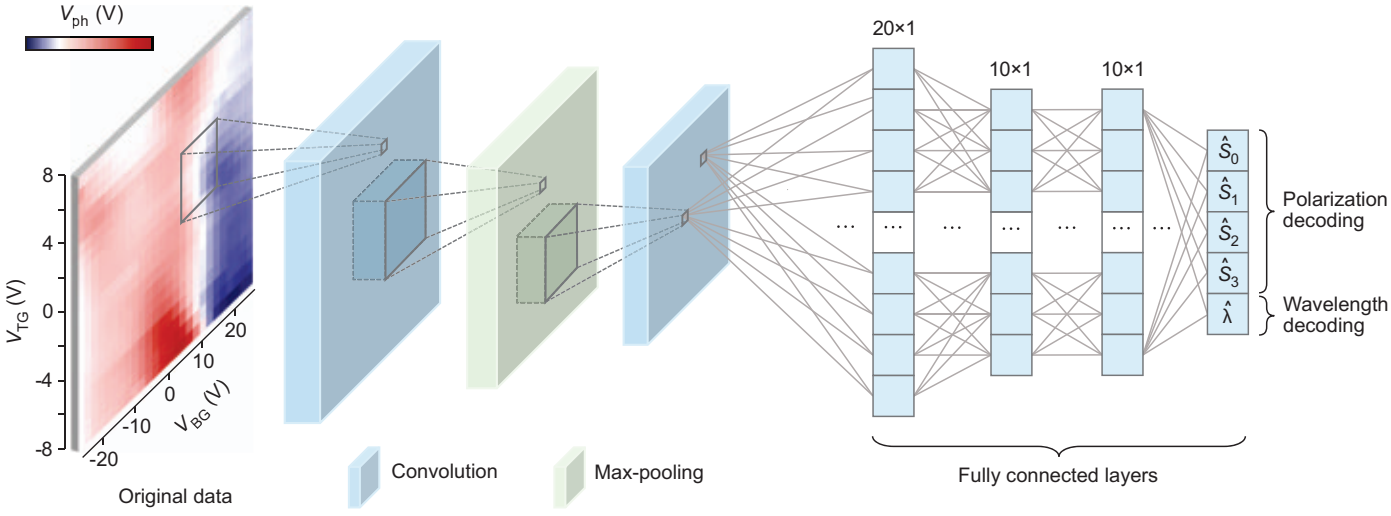


Fig. 5. Deep neural network polarimetry and wavelength detection. Schematic of the convolutional neural network used for the demonstration of the polarimeter and wavelength detection. The input layer is the measured 20-pixel–by–26-pixel photovoltage (V_{ph}) mapping (leftmost panel), and the output is a five-element vector, $\left(\hat{S}_0, \hat{S}_1, \hat{S}_2, \hat{S}_3, \hat{\lambda}\right)$. The mapping, in which the polarization and wavelength information of the incident 5-μm light is encoded, consists of 20 pixels by 26 pixels, which corresponds to 20 values of top-gate voltage (V_{TG}) and 26 values of back-gate voltage (V_{BG}). The hidden layers consist of the first convolution layer, the max-pooling layer, the second convolution layer, and three subsequent fully connected layers. Stokes parameters can be directly calculated using $\left(\hat{S}_0, \hat{S}_1, \hat{S}_2, \hat{S}_3\right)$. A wavelength label $\hat{\lambda} = \mp 1$ corresponds to 5 and 7.7 μm, respectively. [Adapted by permission from Springer Nature Customer Service Center GmbH, Springer *Nature* (*47*), copyright (2022)]

simultaneously using a reconfigurable sensor. As discussed in the section "Information-encoding mechanisms," by using the BPVE, the polarization state, wavelength, and power information can be encoded into a 2D photoresponse map (*47*), as shown in the leftmost panel of Fig. 5. Although the mechanisms of the BPVE are well understood within the framework of quantum geometry, precise analytical modeling of the measured photoresponse is not feasible because of the extrinsic complexities of the moiré system, such as finite temperature, unintentional strain, and twist-angle disorder (*67*). Instead, a convolutional neural network can be trained as the decoder by using a large number of such 2D mappings from excitation lights with known physical properties. The trained convolutional neural network can then be used to decipher the 2D mapping to reveal the wavelength, power, and polarization state of an unknown light (*47*), as illustrated in Fig. 5. We expect deep neural networks to play an increasingly important role as the sensing tasks become more complex and demanding.

Discussion and outlook

Emerging opportunities in deep optical sensing

As noted above, a single TDBG sensor can simultaneously detect the wavelength, intensity, and polarization state of unknown light in the mid-infrared regime. Extending this capability to other wavelength ranges and enabling the detection of other physical properties, such as angular momentum (*88*–*91*), will further empower this sensing scheme. Innovative materials will be needed to demonstrate deep optical sensing beyond the mid-infrared. Moreover, the tunable BPVE, which is central to the TDBG reconfigurable sensor concept, is a second-order optical effect (*47*). The regular photovoltaic and photoconductive responses can also be reconfigurable (*37*–*42*). The use of reconfigurable linear and higher-order photoresponses together may further enhance the capabilities and improve the performance of deep optical sensing.

Another future direction is to expand the capability of functionality multiplexing. It has been shown that an array of reconfigurable photodetectors can realize image-encoding and -classification functions based on artificial neural networks (*37*). However, the device can only handle simple images because it is limited by the low resolution of the array and the low complexity of the neural network architecture. The construction of reconfigurable sensor arrays with higher resolution and more layers will enable the use of deep neural networks in conjunction with enhanced imaging capabilities for more challenging machine-vision tasks. Simultaneous encoding of both spectral and spatial information may lead to a new generation of high-throughput hyperspectral imaging systems.

Identifying innovative sensing materials and mechanisms

As sensors become ubiquitous, it is highly desirable to continuously reduce their size to enable on-chip integration. To achieve this goal, reconfigurability is key, as emphasized throughout this Review. Miniaturized sensors that can perform a wide range of different tasks have been demonstrated mostly using 2D materials such as black phosphorus (*25*), transition metal dichalcogenides (*37*), moiré graphene (*47*), and perovskites (*27*). Research on other material systems to realize reconfigurability will likely extend the operational spectral range and enable new functionalities. Van der Waals heterostructures, for example, represent a diverse spectrum of material systems with strong tunability that can interact with light in the wavelength range from microwave to ultraviolet. External electric fields can tune not only the doping (or Fermi energy) and the bandgap of the constituent materials in the heterostructure but also the relative band alignments between different layers (*26*) and the quantum geometric properties (*47*, *57*). As a result, optical transitions within and between the layers, as well as nonlinear optical effects, can all be reconfigured by electric fields, which provides ample opportunities for the realization of deep optical sensing in a broad wavelength range. Moreover, conventional thin-film semiconductors such as silicon-germanium and III-V quantum wells also exhibit tunability under electric fields (*72*, *73*, *92*), which makes it feasible to build reconfigurable sensors based on highly mature semiconductor platforms. A recently demonstrated silicon reconfigurable imager has shown the potential of using silicon simultaneously for imaging and in-sensor data processing in the visible spectral range (*93*). Compressive sensing and imaging can also benefit from device reconfigurability (*46*, *65*).

Integration of sensing and computing functionalities

Reconfigurable sensors may generate more data than conventional sensors because of their multiple operational states. Reading and processing these large amounts of data can be challenging. Developing information processing schemes with integrated memory cells near or within the sensors may thus be crucial. For example, a reconfigurable integrated sensor array, based on a van der Waals heterostructure, has been developed that incorporates sensing, memory, and computing functions (*41*). This sensor array exhibits nonvolatile negative and positive photoresponses, which are used for motion detection, and external neural networks can be combined with the sensor to enable more advanced functionalities. A recent article (*94*) has extensively discussed the possible strategies for performing in- or near-sensor computing. Deep optical sensing is expected to benefit tremendously from this related field. Further, it is expected that the hybrid integration of reconfigurable sensors made from 2D materials, perovskites, and other thin-film semiconductors with silicon electronics for data processing may pave the way for a new generation of deep sensing technologies.

Optical computing is another promising pathway for high-throughput information processing, which benefits both encoders and decoders. Optical computing can enable higher degrees of freedom in encoder designs by directly processing the optical signals and high-throughput decoding through deep neural networks in hardware format, owing to the high bandwidth of optical modulation and high speed of light. Functions such as computer vision and language processing have been demonstrated based on optical deep neural networks (*95*, *96*).

Leveraging the latest developments in machine learning

The development of deep optical sensing schemes with reconfigurable sensors provides a singular opportunity to test and exploit the latest developments in machine learning. For example, a generative-adversarial network (GAN) has been used to enable compressed sensing without assuming sparsity (*97*). A GAN typically consists of two competing neural networks: a generator and a discriminator (*49*). The generator is trained to produce vectors from random noise to fool the discriminator, whereas the discriminator attempts to distinguish the generated vectors from existing datasets. By properly training these two neural networks, the generator learns the distribution of existing data and eventually generates vectors that possess the features of the dataset. By incorporating prior experience or knowledge, the reconstruction error can be made small even with a limited number of measurements. We expect that GANs can be applied to improve the deep sensing performance in terms of the reduction of required measurements. Another example is the long short-term memory (LSTM), which can be used together with reconfigurable optical sensors to detect ultrafast events in computer vision and chemical reactions (*49*).

Establishing mathematical guidelines for encoding

One challenge in information encoding is to minimize the required number of measurement states m and to identify an optimal encoding strategy. If fewer measurements can be performed without compromising sensing performance, then not only can the acquisition speed be increased, but the data processing requirements in subsequent steps can also be reduced. By contrast, by choosing a large m, redundancy can be introduced, which reduces the probability of degeneracy of the photoresponse. Therefore, a set of mathematical guidelines are needed to bridge the gap between these two conflicting requirements.

In conclusion, optical sensing will benefit tremendously from the latest developments in device technology, materials science, condensed matter physics, and machine learning. Future sensors are likely to be highly compact, reconfigurable, multifunctional, and intelligent, and they will find applications in medical imaging, environmental monitoring, infrared astronomy, and many other areas of our daily lives, especially in the mobile domain.

REFERENCES AND NOTES

1. D. C. D. Pocock, Sight and knowledge. *Trans. Inst. Br. Geogr.* **6**, 385–393 (1981). doi: 10.2307/621875
2. Z. Yang, T. Albrow-Owen, W. Cai, T. Hasan, Miniaturization of optical spectrometers. *Science* **371**, eabe0722 (2021). doi: 10.1126/science.abe0722; pmid: 33509998
3. L. Gao, Y. Qu, L. Wang, Z. Yu, Computational spectrometers enabled by nanophotonics and deep learning. *Nanophotonics* **11**, 2507–2529 (2022). doi: 10.1515/nanoph-2021-0636
4. T. Yang *et al*., Miniature spectrometer based on diffraction in a dispersive hole array. *Opt. Lett.* **40**, 3217–3220 (2015). doi: 10.1364/OL.40.003217; pmid: 26125406
5. B. Redding, S. Fatt Liew, Y. Bromberg, R. Sarma, H. Cao, Evanescently coupled multimode spiral spectrometer. *Optica* **3**, 956–962 (2016). doi: 10.1364/OPTICA.3.000956
6. P. Edwards *et al*., Smartphone based optical spectrometer for diffusive reflectance spectroscopic measurement of hemoglobin. *Sci. Rep.* **7**, 12224 (2017). doi: 10.1038/s41598-017-12482-5; pmid: 28939898
7. M. Faraji-Dana *et al*., Compact folded metasurface spectrometer. *Nat. Commun.* **9**, 4196 (2018). doi: 10.1038/s41467-018-06495-5; pmid: 30305616
8. W. Hartmann *et al*., Waveguide-integrated broadband spectrometer based on tailored disorder. *Adv. Opt. Mater.* **8**, 1901602 (2020). doi: 10.1002/adom.201901602
9. A. V. Velasco *et al*., High-resolution Fourier-transform spectrometer chip with microphotonic silicon spiral waveguides. *Opt. Lett.* **38**, 706–708 (2013). doi: 10.1364/OL.38.000706; pmid: 23455272
10. D. M. Kita *et al*., High-performance and scalable on-chip digital Fourier transform spectroscopy. *Nat. Commun.* **9**, 4405 (2018). doi: 10.1038/s41467-018-06773-2; pmid: 30353014
11. S. N. Zheng *et al*., Microring resonator-assisted Fourier transform spectrometer with enhanced resolution and large bandwidth in single chip solution. *Nat. Commun.* **10**, 2349 (2019). doi: 10.1038/s41467-019-10282-1; pmid: 31138800
12. D. Pohl *et al*., An integrated broadband spectrometer on thin-film lithium niobate. *Nat. Photonics* **14**, 24–29 (2020). doi: 10.1038/s41566-019-0529-9
13. J. Bao, M. G. Bawendi, A colloidal quantum dot spectrometer. *Nature* **523**, 67–70 (2015). doi: 10.1038/nature14576; pmid: 26135449
14. E. Huang, Q. Ma, Z. Liu, Etalon array reconstructive spectrometry. *Sci. Rep.* **7**, 40693 (2017). pmid: 28074883
15. B. Craig, V. R. Shrestha, J. Meng, J. J. Cadusch, K. B. Crozier, Experimental demonstration of infrared spectral reconstruction using plasmonic metasurfaces. *Opt. Lett.* **43**, 4481–4484 (2018). doi: 10.1364/OL.43.004481; pmid: 30211895
16. A. Tittl *et al*., Imaging-based molecular barcoding with pixelated dielectric metasurfaces. *Science* **360**, 1105–1109 (2018). doi: 10.1126/science.aas9768; pmid: 29880685
17. Y. Zhu, X. Lei, K. X. Wang, Z. Yu, Compact CMOS spectral sensor for the visible spectrum. *Photon. Res.* **7**, 961–966 (2019). doi: 10.1364/PRJ.7.000961
18. J. Meng, J. J. Cadusch, K. B. Crozier, Detector-only spectrometer based on structurally colored silicon nanowires and a reconstruction algorithm. *Nano Lett.* **20**, 320–328 (2020). doi: 10.1021/acs.nanolett.9b03862; pmid: 31829611
19. Z. Wang *et al*., Single-shot on-chip spectral sensors based on photonic crystal slabs. *Nat. Commun.* **10**, 1020 (2019). doi: 10.1038/s41467-019-08994-5; pmid: 30833569
20. Z. Yang *et al*., Single-nanowire spectrometers. *Science* **365**, 1017–1020 (2019). doi: 10.1126/science.aax8814; pmid: 31488686
21. X. Zhu *et al*., Broadband perovskite quantum dot spectrometer beyond human visual resolution. *Light Sci. Appl.* **9**, 73 (2020). doi: 10.1038/s41377-020-0301-4; pmid: 32377335
22. J. Zhang, X. Zhu, J. Bao, Denoising autoencoder aided spectrum reconstruction for colloidal quantum dot spectrometers. *IEEE Sens. J.* **21**, 6450–6458 (2020). doi: 10.1109/JSEN.2020.3039973
23. C. Brown *et al*., Neural network-based on-chip spectroscopy using a scalable plasmonic encoder. *ACS Nano* **15**, 6305–6315 (2021). doi: 10.1021/acsnano.1c00079; pmid: 33543919
24. K. D. Hakkel *et al*., Integrated near-infrared spectral sensing. *Nat. Commun.* **13**, 103 (2022). doi: 10.1038/s41467-021-27662-1; pmid: 35013200
25. S. Yuan, D. Naveh, K. Watanabe, T. Taniguchi, F. Xia, A wavelength-scale black phosphorus spectrometer. *Nat. Photonics* **15**, 601–607 (2021). doi: 10.1038/s41566-021-00787-x
26. W. Deng *et al*., Electrically tunable two-dimensional heterojunctions for miniaturized near-infrared spectrometers. *Nat. Commun.* **13**, 4627 (2022). doi: 10.1038/s41467-022-32306-z; pmid: 35941126
27. L. Guo *et al*., A single-dot perovskite spectrometer. *Adv. Mater.* **34**, e2200221 (2022). doi: 10.1002/adma.202200221; pmid: 35706366
28. H. H. Yoon *et al*., Miniaturized spectrometers with a tunable van der Waals junction. *Science* **378**, 296–299 (2022). doi: 10.1126/science.add8544; pmid: 36264793
29. L. Tong *et al*., Stable mid-infrared polarization imaging based on quasi-2D tellurium at room temperature. *Nat. Commun.* **11**, 2308 (2020). doi: 10.1038/s41467-020-16125-8; pmid: 32385242
30. J. Wei, C. Xu, B. Dong, C.-W. Qiu, C. Lee, Mid-infrared semimetal polarization detectors with configurable polarity transition. *Nat. Photonics* **15**, 614–621 (2021). doi: 10.1038/s41566-021-00819-6
31. J. Wei *et al*., Geometric filterless photodetectors for mid-infrared spin light. *Nat. Photonics* **17**, 171–178 (2022). doi: 10.1038/s41566-022-01115-7
32. F. Yesilkoy *et al*., Ultrasensitive hyperspectral imaging and biodetection enabled by dielectric metasurfaces. *Nat. Photonics* **13**, 390–396 (2019). doi: 10.1038/s41566-019-0394-6
33. K. Monakhova, K. Yanny, N. Aggarwal, L. Waller, Spectral diffusercam: Lensless snapshot hyperspectral imaging with a spectral filter array. *Optica* **7**, 1298–1307 (2020). doi: 10.1364/OPTICA.397214
34. W. Zhang *et al*., Deeply learned broadband encoding stochastic hyperspectral imaging. *Light Sci. Appl.* **10**, 108 (2021). doi: 10.1038/s41377-021-00545-2; pmid: 34035213
35. Z. Ballard, C. Brown, A. M. Madni, A. Ozcan, Machine learning and computation-enabled intelligent sensor design. *Nat. Mach. Intell.* **3**, 556–565 (2021). doi: 10.1038/s42256-021-00360-9
36. Y. Luo *et al*., Design of task-specific optical systems using broadband diffractive neural networks. *Light Sci. Appl.* **8**, 112 (2019). doi: 10.1038/s41377-019-0223-1; pmid: 31814969

37. L. Mennel *et al*., Ultrafast machine vision with 2D material neural network image sensors. *Nature* **579**, 62–66 (2020). doi: 10.1038/s41586-020-2038-x; pmid: 32132692
38. H. Jang *et al*., An atomically thin optoelectronic machine vision processor. *Adv. Mater.* **32**, e2002431 (2020). doi: 10.1002/adma.202002431; pmid: 32700395
39. C.-Y. Wang *et al*., Gate-tunable van der Waals heterostructure for reconfigurable neural network vision sensor. *Sci. Adv.* **6**, eaba6173 (2020). doi: 10.1126/sciadv.aba6173; pmid: 32637614
40. T. Ahmed *et al*., Fully light-controlled memory and neuromorphic computation in layered black phosphorus. *Adv. Mater.* **33**, e2004207 (2021). doi: 10.1002/adma.202004207; pmid: 33205523
41. Z. Zhang *et al*., All-in-one two-dimensional retinomorphic hardware device for motion detection and recognition. *Nat. Nanotechnol.* **17**, 27–32 (2022). doi: 10.1038/s41565-021-01003-1; pmid: 34750561
42. S. Lee, R. Peng, C. Wu, M. Li, Programmable black phosphorus image sensor for broadband optoelectronic edge computing. *Nat. Commun.* **13**, 1485 (2022). doi: 10.1038/s41467-022-29171-1; pmid: 35304489
43. S. Wang *et al*., Networking retinomorphic sensor with memristive crossbar for brain-inspired visual perception. *Natl. Sci. Rev.* **8**, nwaa172 (2020). doi: 10.1093/nsr/nwaa172; pmid: 34691573
44. Q.-B. Zhu *et al*., A flexible ultrasensitive optoelectronic sensor array for neuromorphic vision systems. *Nat. Commun.* **12**, 1798 (2021). doi: 10.1038/s41467-021-22047-w; pmid: 33741964
45. C. Choi *et al*., Curved neuromorphic image sensor array using a MoS_2-organic heterostructure inspired by the human visual recognition system. *Nat. Commun.* **11**, 5934 (2020). doi: 10.1038/s41467-020-19806-6
46. L. Mennel, D. K. Polyushkin, D. Kwak, T. Mueller, Sparse pixel image sensor. *Sci. Rep.* **12**, 5650 (2022). doi: 10.1038/s41598-022-09594-y; pmid: 35383216
47. C. Ma *et al*., Intelligent infrared sensing enabled by tunable moiré quantum geometry. *Nature* **604**, 266–272 (2022). doi: 10.1038/s41586-022-04548-w; pmid: 35418636
48. M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, P. Vandergheynst, Geometric deep learning: Going beyond euclidean data. *IEEE Signal Process. Mag.* **34**, 18–42 (2017). doi: 10.1109/MSP.2017.2693418
49. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning* (MIT Press, 2016).
50. Q. Ma, A. G. Grushin, K. S. Burch, Topology and geometry under the nonlinear electromagnetic spotlight. *Nat. Mater.* **20**, 1601–1614 (2021). doi: 10.1038/s41563-021-00992-7; pmid: 34127824
51. A. M. Cook, B. M Fregoso, F. de Juan, S. Coh, J. E. Moore, Design principles for shift current photovoltaics. *Nat. Commun.* **8**, 14176 (2017). doi: 10.1038/ncomms14176; pmid: 28120823
52. T. Morimoto, N. Nagaosa, Topological nature of nonlinear optical effects in solids. *Sci. Adv.* **2**, e1501524 (2016). doi: 10.1126/sciadv.1501524; pmid: 27386523
53. J. Sipe, A. Shkrebtii, Second-order optical response in semiconductors. *Phys. Rev. B* **61**, 5337–5352 (2000). doi: 10.1103/PhysRevB.61.5337
54. J. Ahn, G.-Y. Guo, N. Nagaosa, A. Vishwanath, Riemannian geometry of resonant optical responses. *Nat. Phys.* **18**, 290–295 (2022). doi: 10.1038/s41567-021-01465-z
55. E. Cohen *et al*., Geometric phase from Aharonov–Bohm to Pancharatnam–Berry and beyond. *Nat. Rev. Phys.* **1**, 437–449 (2019). doi: 10.1038/s42254-019-0071-1
56. Q. Ma *et al*., Direct optical detection of Weyl fermion chirality in a topological semimetal. *Nat. Phys.* **13**, 842–847 (2017). doi: 10.1038/nphys4146
57. S.-Y. Xu *et al*., Electrically switchable Berry curvature dipole in the monolayer topological insulator WTe_2. *Nat. Phys.* **14**, 900–906 (2018). doi: 10.1038/s41567-018-0189-6
58. G. B. Osterhoudt *et al*., Colossal mid-infrared bulk photovoltaic effect in a type-I Weyl semimetal. *Nat. Mater.* **18**, 471–475 (2019). doi: 10.1038/s41563-019-0297-4; pmid: 30833781
59. J. Ma *et al*., Nonlinear photoresponse of type-II Weyl semimetals. *Nat. Mater.* **18**, 476–481 (2019). doi: 10.1038/s41563-019-0296-5; pmid: 30833780
60. T. Akamatsu *et al*., A van der Waals interface that creates in-plane polarization and a spontaneous photovoltaic effect. *Science* **372**, 68–72 (2021). doi: 10.1126/science.aaz9146; pmid: 33795452
61. T. M. Cover, *Elements of Information Theory* (Wiley, 1999).
62. Y. August, A. Stern, Compressive sensing spectrometry based on liquid crystal devices. *Opt. Lett.* **38**, 4996–4999 (2013). doi: 10.1364/OL.38.004996; pmid: 24281493
63. D. L. Donoho, Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006). doi: 10.1109/TIT.2006.871582
64. Y. C. Eldar, G. Kutyniok, *Compressed Sensing: Theory and Applications* (Cambridge Univ. Press, 2012).
65. M. F. Duarte *et al*., Single-pixel imaging via compressive sampling. *IEEE Signal Process. Mag.* **25**, 83–91 (2008). doi: 10.1109/MSP.2007.914730
66. N. A. Rubin *et al*., Matrix Fourier optics enables a compact full-Stokes polarization camera. *Science* **365**, eaax1839 (2019). doi: 10.1126/science.aax1839; pmid: 31273096
67. C. N. Lau, M. W. Bockrath, K. F. Mak, F. Zhang, Reproducibility in the fabrication and physics of moiré materials. *Nature* **602**, 41–50 (2022). doi: 10.1038/s41586-021-04173-z; pmid: 35110759
68. J. Xiong, X. Cai, K. Cui, Y. Huang, J. Yang, H. Zhu, Z. Zheng, S. Xu, Y. He, F. Liu, X. Feng, W. Zhang, One-shot ultraspectral imaging with reconfigurable metasurfaces. arXiv:2005.02689 [physics.optics] (2020).
69. Z. Wang, Z. Yu, Spectral analysis based on compressive sensing in nanophotonic structures. *Opt. Express* **22**, 25608–25614 (2014). doi: 10.1364/OE.22.025608; pmid: 25401594
70. J. J. Cadusch, J. Meng, B. Craig, K. B. Crozier, Silicon microspectrometer chip based on nanostructured fishnet photodetectors with tailored responsivities and machine learning. *Optica* **6**, 1171–1177 (2019). doi: 10.1364/OPTICA.6.001171
71. S. L. Chuang, *Physics of Photonic Devices* (Wiley, 2012).
72. D. A. Miller *et al*., Band-edge electroabsorption in quantum well structures: The quantum-confined stark effect. *Phys. Rev. Lett.* **53**, 2173–2176 (1984). doi: 10.1103/PhysRevLett.53.2173
73. Y.-H. Kuo *et al*., Strong quantum-confined Stark effect in germanium quantum-well structures on silicon. *Nature* **437**, 1334–1336 (2005). doi: 10.1038/nature04204; pmid: 16251959
74. S. A. Empedocles, M. G. Bawendi, Quantum-confined stark effect in single CdSe nanocrystallite quantum dots. *Science* **278**, 2114–2117 (1997). doi: 10.1126/science.278.5346.2114; pmid: 9405345
75. C. Lin, R. Grassi, T. Low, A. S. Helmy, Multilayer black phosphorus as a versatile mid-infrared electro-optic material. *Nano Lett.* **16**, 1683–1689 (2016). doi: 10.1021/acs.nanolett.5b04594; pmid: 26901350
76. Y. Yao *et al*., Electrically tunable metasurface perfect absorbers for ultrathin mid-infrared optical modulators. *Nano Lett.* **14**, 6526–6532 (2014). doi: 10.1021/nl503104n; pmid: 25310847
77. Y. Yao *et al*., Broad electrical tuning of graphene-loaded plasmonic antennas. *Nano Lett.* **13**, 1257–1264 (2013). doi: 10.1021/nl3047943; pmid: 23441688
78. V. R. Shrestha *et al*., Mid- to long-wave infrared computational spectroscopy with a graphene metasurface modulator. *Sci. Rep.* **10**, 5377 (2020). doi: 10.1038/s41598-020-61998-w; pmid: 32214114
79. L. Mennel *et al*., A photosensor employing data-driven binning for ultrafast image recognition. *Sci. Rep.* **12**, 14441 (2022). doi: 10.1038/s41598-022-18821-5; pmid: 36002539
80. D. Ulyanov, A. Vedaldi, V. Lempitsky, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2018), pp. 9446–9454.
81. G. Mataev, P. Milanfar, M. Elad, DeepRED: Deep image prior powered by RED. arxiv:1903.10176 [cx.CV] (2019).
82. K. H. Jin, M. T. McCann, E. Froustey, M. Unser, Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Image Process.* **26**, 4509–4522 (2017). doi: 10.1109/TIP.2017.2713099; pmid: 28641250
83. A. N. Tikhonov, A. Goncharsky, V. Stepanov, A. G. Yagola, *Numerical Methods for the Solution of Ill-Posed Problems* (Springer, 1995), vol. 328.
84. R. Tibshirani, Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Series B Stat. Methodol.* **58**, 267–288 (1996).
85. C. Shorten, T. M. Khoshgoftaar, A survey on image data augmentation for deep learning. *J. Big Data* **6**, 60 (2019). doi: 10.1186/s40537-019-0197-0
86. M. Prezioso *et al*., Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* **521**, 61–64 (2015). doi: 10.1038/nature14441; pmid: 25951284
87. C. Li *et al*., Analogue signal and image processing with large memristor crossbars. *Nat. Electron.* **1**, 52–59 (2018). doi: 10.1038/s41928-017-0002-z
88. D. L. Andrews, M. Babiker, *The Angular Momentum of Light* (Cambridge Univ. Press, 2012).
89. L. Allen, M. W. Beijersbergen, R. J. Spreeuw, J. P. Woerdman, Orbital angular momentum of light and the transformation of Laguerre-Gaussian laser modes. *Phys. Rev. A* **45**, 8185–8189 (1992). doi: 10.1103/PhysRevA.45.8185; pmid: 9906912
90. A. M. Yao, M. J. Padgett, Orbital angular momentum: Origins, behavior and applications. *Adv. Opt. Photonics* **3**, 161–204 (2011). doi: 10.1364/AOP.3.000161
91. Z. Ji *et al*., Photocurrent detection of the orbital angular momentum of light. *Science* **368**, 763–767 (2020). doi: 10.1126/science.aba9192; pmid: 32409474
92. R. Karunasiri, Y. Mii, K. L. Wang, Tunable infrared modulator and switch using stark shift in step quantum wells. *IEEE Electron Device Lett.* **11**, 227–229 (1990). doi: 10.1109/55.55258
93. H. Jang *et al*., In-sensor optoelectronic computing using electrostatically doped silicon. *Nat. Electron.* **5**, 519–525 (2022). doi: 10.1038/s41928-022-00819-6
94. F. Zhou, Y. Chai, Near-sensor and in-sensor computing. *Nat. Electron.* **3**, 664–671 (2020). doi: 10.1038/s41928-020-00501-9
95. X. Lin *et al*., All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018). doi: 10.1126/science.aat8084; pmid: 30049787
96. Y. Shen *et al*., Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **11**, 441–446 (2017). doi: 10.1038/nphoton.2017.93
97. A. Bora, A. Jalal, E. Price, A. G. Dimakis, in *Proceedings of the 34th International Conference on Machine Learning* (Proceedings of Machine Learning Research, 2017), pp. 537–546.

ACKNOWLEDGMENTS

Funding: T.M. acknowledges support from the European Union (grant agreement no. 785219 Graphene Flagship). F.Z. acknowledges support from the Army Research Office under grant number W911NF-18-1-0416 and the National Science Foundation (NSF) under grant numbers DMR-1945351 through the Faculty Early Career Development Program (CAREER), DMR-2105139 through the Condensed Matter Physics (CMP) program, and DMR-1921581 through the Designing Materials to Revolutionize and Engineer our Future (DMREF) program. F.X., S.Y., and C.M. acknowledge support from the NSF Emerging Frontiers in Research and Innovation (EFRI) NewLAW program under grant number 1741693, NSF grant 2150561, the Yale Raymond John Wean Foundation, and the government of Israel. D.N. acknowledges support from the Binational Science Foundation under Electrical, Communications, and Cyber Systems (ECCS) grant NSF-BSF 2021721. **Competing interests:** S.Y., C.M., F.Z., and F.X. have a provisional patent application on intelligent light sensing. S.Y., D.N., and F.X. have a pending patent application on on-chip spectroscopy, and D.N. and F.X. have a provisional patent application on the use of on-chip spectroscopy for automobile safety monitoring. T.M. has a pending patent application on ultrafast machine vision. At present, S.Y. is affiliated with KLA Corporation in Milpitas, CA, USA; all his contributions were made based on his work at Yale University.

Submitted 17 October 2022; accepted 15 February 2023
10.1126/science.ade1220

RESEARCH ARTICLE SUMMARY

IMMUNOLOGY

Cysteine carboxyethylation generates neoantigens to induce HLA-restricted autoimmunity

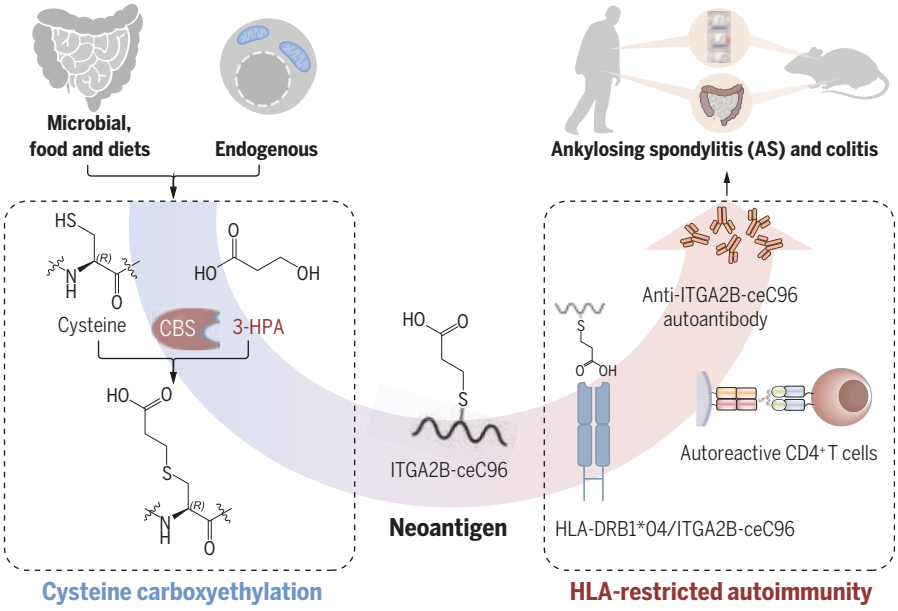
Yue Zhai†, Liang Chen*†, Qian Zhao†, Zhao-Hui Zheng†, Zhi-Nan Chen†, Huijie Bian†, Xu Yang‡, Huan-Yu Lu‡, Peng Lin‡, Xi Chen‡, Ruo Chen‡, Hao-Yang Sun, Lin-Ni Fan, Kun Zhang, Bin Wang, Xiu-Xuan Sun, Zhuan Feng, Yu-Meng Zhu, Jian-Sheng Zhou, Shi-Rui Chen, Tao Zhang, Si-Yu Chen, Jun-Jie Chen, Kui Zhang, Yan Wang, Yang Chang, Rui Zhang, Bei Zhang, Li-Juan Wang, Xiao-Min Li, Qian He, Xiang-Min Yang, Gang Nan, Rong-Hua Xie, Liu Yang, Jing-Hua Yang*, Ping Zhu*

INTRODUCTION: Autoimmune diseases such as ankylosing spondylitis (AS) can be caused by emerging neoantigens that break immune tolerance in humans. Posttranslational modifications (PTMs) have been shown to be a critical mechanism that alters protein structure and function to generate neoantigens and induce subsequent autoimmune responses. Previous studies have confirmed that citrulline-modified peptides are a critical source of neoantigens in rheumatoid arthritis. However, the molecular mechanisms underlying neoantigen formation and pathogenic autoreactive responses for AS are largely unknown. There is an urgent need to develop a systematic approach to profiling the possible PTMs in patients with AS and identifying AS-associated PTMs responsible for autoreactive neoantigen production to better understand the etiology of autoimmune diseases.

RATIONALE: AS has been suggested to be an autoimmune disease because of its clear correlation with certain major histocompatibility complex (MHC) alleles, including HLA-B27. Neoantigens have been hypothesized to induce an aberrant immune response, leading to pathogenic autoreactive T cell responses and autoantibody generation in AS. Here, we developed a systematic open search approach to identify any possible amino acid residues and derivatives in the proteins that are different from the genomic coding sequences. We then applied this information to identify AS-related neoantigens with PTMs within a possible pool of PTM autoantigens and elucidate the pathogenesis of AS.

RESULTS: An open search approach was applied to identify any possible amino acid derivatives across the proteome of patients with AS. This approach generated a large set of noncoded amino acids representing the mass differences between the coded amino acids and actual residues. Among these, an amino acid derivative with a delta mass of 72.021 showed the greatest increase in patients with AS and resulted from a PTM called cysteine carboxyethylation. In vitro and in vivo experiments demonstrated that carboxyethylation at a cysteine residue of integrin αIIb [ITGA2B (CD41)] was catalyzed by cystathionine beta synthase (CBS) in a process that required 3-hydroxypropionic acid (3-HPA), a metabolite commonly released from gut microbes. Cysteine carboxyethylation induced the lysosomal degradation of ITGA2B and produced neoantigens that triggered MHC-II–dependent $CD4^+$ T cell responses. Fluorescence polarization and enzyme-linked immunosorbent assay (ELISA) demonstrated that the identified carboxyethylated peptide (ITGA2B-ceC96) specifically interacted with HLA-DRA*01/HLA-DRB1*04 and was associated with autoantibody production and T cell responses in HLA-DRB1*04 patients. Additional in vitro assays showed that the neoantigen ITGA2B-ceC96 correlated with 3-HPA levels but was independent of CBS expression. HLA-DRB1 haplotype, the carboxyethylated peptide, specific autoantibodies, and 3-HPA levels in patients with AS all correlated with one another. 3-HPA–treated and ITGA2B-ceC96–immunized HLA-DR4 transgenic mice developed colitis and vertebral bone erosion. Thus, cysteine carboxyethylation induced by the metabolite 3-HPA generates a neoantigen that appears to be critical for autoimmune responses in patients with AS.

CONCLUSION: Cysteine carboxyethylation is an in vivo protein modification induced by the metabolite 3-HPA, which is commonly released from gut microbes. Carboxyethylated ITGA2B then induces autoantibody production and autoimmune response in AS. Our work provides a systematic workflow to identify differentially modified proteins that are important for neoantigen production in immune disorders. This approach furthers our understanding of AS pathogenesis and may aid in the development of neoantigen-based diagnosis and treatment for AS and other autoimmune diseases. ■



Metabolite-induced cysteine carboxyethylation provokes HLA-restricted autoimmune responses in ankylosing spondylitis. 3-HPA, which is commonly obtained from food and gut microbes, induces carboxyethylation of cysteine residues in integrin αIIb (ITGA2B). Cysteine carboxyethylation requires CBS, and carboxyethylated ITGA2B (ITGA2B-ceC96) peptides are recruited to the HLA-DR4 complex and thereby stimulate $CD4^+$ T cell responses closely related to AS.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: zhuping@fmmu.edu.cn (P.Z.); jhy@zzu.edu.cn (J.-H.Y.); lchen1@shu.edu.cn (L.C.)
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
Cite this article as Y. Zhai *et al.*, *Science* **379**, eabg2482 (2023). DOI: 10.1126/science.abg2482

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abg2482

RESEARCH ARTICLE

IMMUNOLOGY

Cysteine carboxyethylation generates neoantigens to induce HLA-restricted autoimmunity

Yue Zhai[1]†, Liang Chen[2]*†, Qian Zhao[3]†, Zhao-Hui Zheng[1]†, Zhi-Nan Chen[1]†, Huijie Bian[1]†, Xu Yang[1]‡, Huan-Yu Lu[4]‡, Peng Lin[1]‡, Xi Chen[1]‡, Ruo Chen[1]‡, Hao-Yang Sun[1], Lin-Ni Fan[5], Kun Zhang[1], Bin Wang[1], Xiu-Xuan Sun[1], Zhuan Feng[1], Yu-Meng Zhu[1], Jian-Sheng Zhou[1], Shi-Rui Chen[1], Tao Zhang[1], Si-Yu Chen[1], Jun-Jie Chen[1], Kui Zhang[1], Yan Wang[1], Yang Chang[1], Rui Zhang[1], Bei Zhang[1], Li-Juan Wang[1], Xiao-Min Li[1], Qian He[1], Xiang-Min Yang[1], Gang Nan[1], Rong-Hua Xie[1], Liu Yang[1], Jing-Hua Yang[1,3]*, Ping Zhu[1]*

Autoimmune diseases such as ankylosing spondylitis (AS) can be driven by emerging neoantigens that disrupt immune tolerance. Here, we developed a workflow to profile posttranslational modifications involved in neoantigen formation. Using mass spectrometry, we identified a panel of cysteine residues differentially modified by carboxyethylation that required 3-hydroxypropionic acid to generate neoantigens in patients with AS. The lysosomal degradation of integrin αIIb [ITGA2B (CD41)] carboxyethylated at Cys96 (ITGA2B-ceC96) generated carboxyethylated peptides that were presented by HLA-DRB1*04 to stimulate CD4$^+$ T cell responses and induce autoantibody production. Immunization of HLA-DR4 transgenic mice with the ITGA2B-ceC96 peptide promoted colitis and vertebral bone erosion. Thus, metabolite-induced cysteine carboxyethylation can give rise to pathogenic neoantigens that lead to autoreactive CD4$^+$ T cell responses and autoantibody production in autoimmune diseases.

Posttranslational modifications (PTMs) can generate noncoded amino acid (ncAA) derivatives that are known to change protein structure and function. These modifications potentially produce neoantigens that can drive autoimmune responses (*1*, *2*). Although a few PTMs have been shown to correlate with autoimmune diseases (*3*, *4*), there have been no systematic approaches to profiling the ncAA derivatives critical for the mechanisms of autoreactive neoantigen production and the etiology and pathology of autoimmune diseases. In this study, we developed a workflow to systematically screen PTMs associated with ankylosing spondylitis (AS) and identified a panel of metabolite-induced cysteine carboxyethylation targets. Among these, a differentially carboxyethylated integrin αIIb (ITGA2B) at Cys96, referred to hereafter as ITGA2B-ceC96, was demonstrated to generate the pathogenic neoantigen presented by HLA-DRB1*04 that activated CD4$^+$ T cells and induced autoimmune responses. Moreover, when transgenic mice bearing the human HLA-DR4 allele were immunized with the ITGA2B-ceC96 peptide, colitis and vertebral bone erosion were induced.

[1]Department of Clinical Immunology, Xijing Hospital, and Department of Cell Biology of National Translational Science Center for Molecular Medicine, Fourth Military Medical University, Xi'an 710032, China. [2]School of Medicine, Shanghai University, Shanghai 200444, China. [3]Clinical Systems Biology Laboratories, Translational Medicine Center, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450001, China. [4]Department of Occupational and Environmental Health and the Ministry of Education Key Lab of Hazard Assessment and Control in Special Operational Environment, School of Public Health, Fourth Military Medical University, Xi'an 710032, China. [5]State Key Laboratory of Cancer Biology, Department of Pathology, Xijing Hospital and School of Basic Medicine, Fourth Military Medical University, Xi'an 710032, China.
***Corresponding author. Email: zhuping@fmmu.edu.cn (P.Z.); jhy@zzu.edu.cn (J.-H.Y.); lchen1@shu.edu.cn (L.C.)**
†These authors contributed equally to this work.
‡These authors contributed equally to this work.

Results

Cysteine carboxyethylation is increased in patients with AS

AS is a prototype of spondyloarthritis that is characterized by inflammation of the spine and sacroiliac joints, leading to aberrant bone erosion, remodeling, and ankylosis (*5*, *6*). AS is additionally characterized by comorbidities including enthesitis and colitis (*7*–*9*). To identify the ncAAs of proteins in peripheral blood mononuclear cells (PBMCs) from patients with AS and healthy donors (table S1), we measured the mass differences between the coded amino acids and the actual residues (data S1). In total, 643 unique delta mass clusters spreading on 21,716 protein positions (table S2) were identified. Most of these matched previously reported PTMs (Fig. 1A) and were present at similar levels in patients with AS and healthy donors (Fig. 1B). However, the delta mass cluster at 72.021 Da was significantly increased in AS patients [72.021 ± 0.007, R^2 = 0.41, fold change (FC) = 2.4, $P < 0.05$, n = 881] (Fig. 1B and table S3). Although this cluster was most commonly detected at Cys, Arg, and Trp residues (Fig. 1C), Cys+72.021 fit best to a normal distribution, exhibited a regression efficiency >0.6 (±0.002, R^2 = 0.64, n = 487) (Fig. 1D and fig. S1A), and was significantly higher in the patients with AS than in the HCs (P = 0.017, total Cys+72.021) (Fig. 1E). Although many proteins were affected by the mass shift of Cys+72.021, ITGA2B was showed the greatest increase in modification in the patient group (FC = 11.71, P = 0.037) (Fig. 1F). Although four different cysteine residues at positions 87, 96, 161, and 718 in ITGA2B were detected by mass shifts (Fig. 1G and fig. S1, B and C), only the Cys96 modification, assigned as ITGA2B 96C+72.021, was significantly increased in patients with AS (P = 0.012) (fig. S1, D and E). Because the cysteine residues at these four positions are known to form disulfide bonds (*10*), the mass shift of ITGA2B 96C+72.021 was expected to disrupt the structure and function of ITGA2B (Fig. 1, H and I) in patients with AS.

The mass shift of 72.021 matched the molecular formula of $C_3H_4O_2$. This result was consistent with either a hydroxypropionyl group or an carboxyethyl group, which are potential products of a condensation reaction between the cysteine thiol group and 3-hydroxypropionate (3-HPA) or lactate (LA). Four possible modifications might be produced through either thioether or thioester bonds, including S-(1-carboxyethyl)-Cys (#1), S-(2-carboxyethyl)-Cys (#3), S-(2-hydroxypropionyl)-Cys (#5), and S-(3-hydroxypropionyl)-Cys (#7) (Fig. 2A). The in vitro reaction of wild-type ITGA2B Cys96 peptide (ITGA2B-wtC96) with 3-HPA and LA led to carboxyethylation (#1) or hydroxypropionylation (#5) (Fig. 2B), indicating that 3-HPA is the likely substrate. Four modified ITGA2B-C96 peptides with isomers #1, #3, #5, and #7 were then synthesized (fig. S2A). The thioester bonds at C96 (#5 and #7) rapidly dissociated (fig. S2, B and C), suggesting that the thioester modifications produced by LA or 3-HPA were very unstable in the aqueous phase and thus unlikely to account for the in vivo formation of a stably modified product. Thus, the Cys+72.021 modification observed in PBMCs is likely the carboxyethylation of cysteine with a thioether bond to 3-HPA.

3-HPA induces the cysteine carboxyethylation of ITGA2B

To confirm the occurrence of cysteine carboxyethylation, we generated a specific antibody against carboxyethylated ITGA2B Cys96 (ITGA2B-ceC96) in rabbits. The anti-ceC96 antibody showed very high affinity and specificity for the carboxyethylated peptide (ceC96, K_D = 5.32×10^{-9}) (Fig. 2, C and D). Indeed, when ITGA2B was transiently expressed in 293T cells in the presence of 3-HPA, the carboxyethylated ITGA2B was clearly detected by immunoblotting using the anti-ceC96 antibody (Fig. 2E). To confirm cysteine carboxyethylation in vivo, endogenous ITGA2B from AS patient PBMCs was immunoprecipitated with the anti-ceC96 antibody (Fig. 2F) and analyzed by liquid chromatography–tandem mass spectometry (LC–MS/MS). The

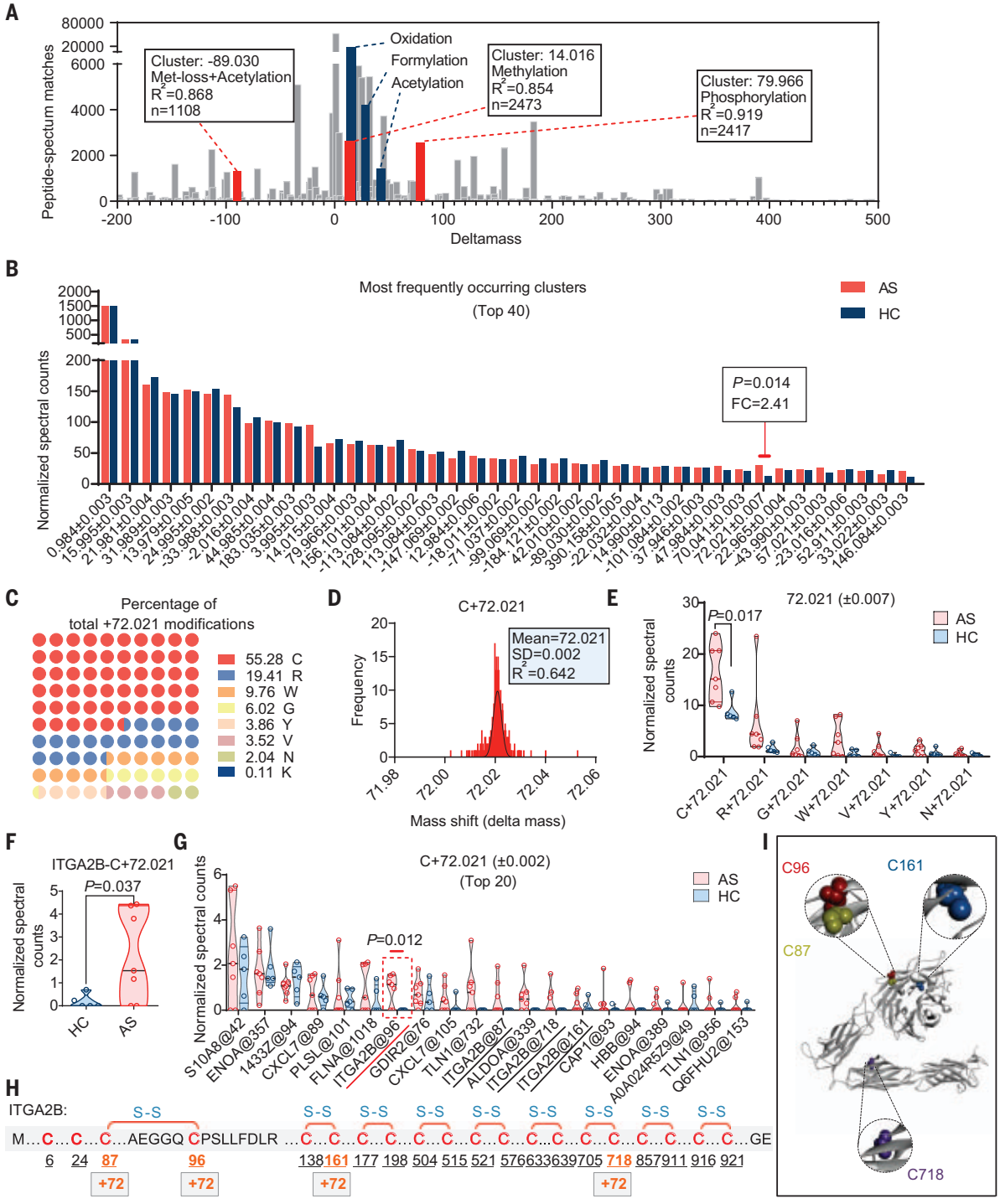


Fig. 1. The cysteine+72.021 modification is significantly enriched in the PBMCs of patients with AS. (**A**) AS patient and HC PBMC mass spectrometry analysis. The delta mass clusters potentially indicate nonstandard amino acids or modified amino acid derivatives. (**B**) The top 40 delta mass clusters with the highest frequencies in the PBMC samples. Among these, one delta mass was 72.021 ± 0.007 Da (for the AS group, n = 31; for HC the group, n = 13) with a fold change (FC) > 2, and the P value was 0.014. Statistical significance was determined using Student's t test. The FDR for 72.021 ± 0.007 was 0.023 according to the Benjamini–Hochberg procedure. These data are representative of three experiments. (**C**) Of the 881 +72.021 ± 0.007 modifications, 55.28% were cysteine (C) modifications (n = 487). (**D**) The frequency distributions of cysteine with delta masses clustered at 72.021. (**E**) The Cys + 72.021 modifications differed between AS patients (n = 7) and HCs (n = 5). Additionally, the number of modifications was higher in the AS patients. Statistical significance was determined using Student's t test. These data are representative of three experiments. (**F**) Mass shift of Cys+72.021 on integrin alpha-IIb (ITGA2B). (**G**) Among the top 20 proteins and sites with the Cys+72.021 modification, there were four Cys +72.021 modifications in the ITGA2B@C96, ITGA2B@C87, ITGA2B@C161, and ITGA2B@C718 proteins, as indicated with red and black lines. The abundance of the ITGA2B@C96 modification was significantly different (P = 0.012) in AS patients (n = 7) and HCs (n = 5). Statistical significance was determined using Student's t test. These data are representative of three experiments. (**H**) ITGA2B contains up to 20 cysteines, of which 18 form disulfide bonds. (**I**) Structural analysis revealed Cys+72.021 modifications on four cysteines within ITGA2B (ITGA2B@C96, ITGA2B@C87, ITGA2B@C161, and ITGA2B@C718) in AS patients (PDB accession no. 1u8c).

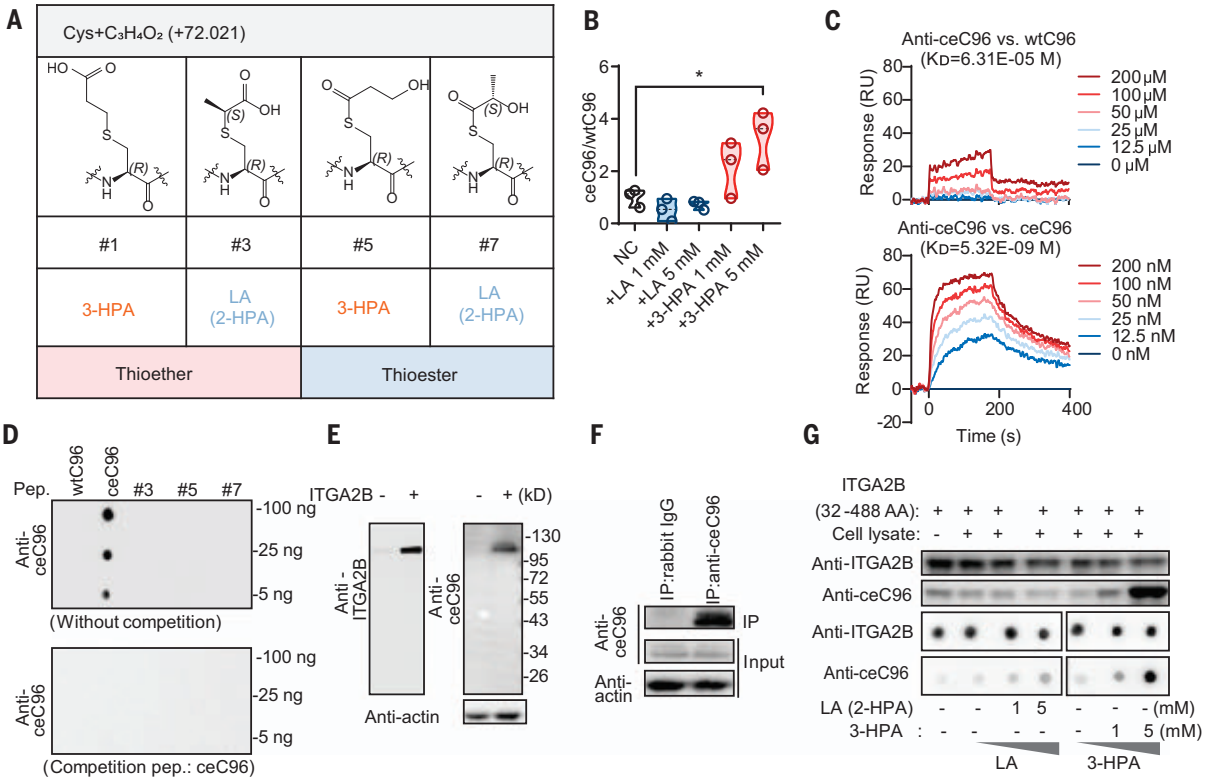


Fig. 2. Cysteine carboxyethylation requires 3-HPA as a substrate. (**A**) The composition of the Cys+72.021 modification is $C_3H_4O_2$. The four primary possible modifications are shown as #1 (ceC96), #3, #5, and #7. The metabolites that are potential substrates for cysteine modification are 3-HPA and LA/2-HPA. (**B**) Modification-related metabolite analysis using LC–MS/MS. The ITGA2B-wtC96(91-104) peptide without modification (AEGGQCPSLLFDLR) was incubated in vitro with either LA/2-HPA or 3-HPA. Statistical significance was determined using one-way ANOVA with Dunnett's multiple-comparisons test with $^*P < 0.05$. $n = 3$ per group per experiment. These data are representative of two experiments. (**C**) SPR analysis of the affinity of the anti-ITGA2B-C96 carboxyethylation antibody (anti-ceC96) for the modified peptide (ceC96) and unmodified peptide (wtC96). The results shown are representative of three independent experiments. (**D**) Dot blot and ceC96 competition assays were used to assess anti-ITGA2B-ceC96 antibody recognition. Unmodified (wtC96), 2-carboxyethyl-modified (ceC96), 1-carboxyethyl-modified (#3), 3-hydroxypropionyl-modified (#5), and 2-hydroxypropionyl-modified (#7) peptides were tested with the anti–ITGA2B-ceC96 (anti-ceC96) antibody. (**E**) Immunoblots of lysates from 293T cells overexpressing ITGA2B and treated with 1 mM 3-HPA. The blots were probed with the anti-ITGA2B-ceC96 and anti-ITGA2B antibodies. (**F**) Patient PBMCs were subjected to endogenous ITGA2B immunoprecipitation with anti-ceC96 antibodies. (**G**) LA/2-HPA and 3-HPA were added to bacterially expressed ITGA2B protein (32 to 488 amino acids) with or without 293T cell lysate. An anti-ITGA2B antibody (recognizing an immunogen from ITGA2B 50 to 100 amino acids) and the polyclonal anti-ceC96 antibody (prepared for ITGA2B-C96 2-carboxyethylation) were used in dot blotting and immunoblotting assays. In (C) to (G), the results shown are representative of three independent experiments.

ITGA2B immunoprecipitated from PBMCs of patients with AS showed a typical MS/MS spectrum similar to that of the synthetic ITGA2B-ceC96 peptide (fig. S2D).

The requirement for 3-HPA in cysteine carboxyethylation was further confirmed in vitro when carboxyethylated ITGA2B appeared and increased in a dose-dependent manner after incubation with 3-HPA (Fig. 2G). Moreover, this reaction was abolished by inactivating cell lysates through heat (fig. S3A). Thus, in vivo cysteine carboxyethylation is an enzymatic reaction that requires the metabolite 3-HPA to provide the necessary carboxyethyl group.

To identify the enzyme that catalyzes cysteine carboxyethylation, protein complexes were immunoprecipitated with specific antibodies against unmodified and modified ITGA2B proteins (anti-ITGA2B and anti–ITGA2B-ceC96, respectively). The components were then identified by LC–MS/MS. Interaction network analysis of the identified components revealed differences in the biological processes between modified versus unmodified ITGA2B. The unmodified ITGA2B complex was associated with vesicle-mediated transport, whereas the ITGA2B-ceC96 complex was associated with the L-cysteine metabolic process, which is linked to cystathionine-β-synthase (CBS) (fig. S3B). CBS was previously shown to convert serine and homocysteine to cystathionine in the folate pathway, requiring pyridoxal-phosphate (PLP) and S-adenosyl-L-methionine (AdoMet) as cofactors and enhancers, respectively (*11*–*13*). We therefore hypothesized that CBS could promote cysteine 2-carboxyethylation (Fig. 3A and fig. S3C). In support of this hypothesis, the carboxyethylation of ITGA2B Cys96 was enhanced by overexpressing CBS. Moreover, PTM levels increased in both a dose- and time-dependent manner (Fig. 3, B and C). The requirement for CBS was further confirmed by an in vitro assay. The cysteine 96 in the synthetic peptide ITGA2B-wtC96(91-104) (Fig. 3D) and the recombinant protein ITGA2B (32-989) (Fig. 3E) were efficiently carboxyethylated by CBS in the presence of 3-HPA and the cofactors PLP and AdoMet.

Cysteine carboxyethylation induces the lysosomal degradation of ITGA2B and produces neoantigens that trigger CD4⁺ T cell responses

We next examined the stability of ITGA2B and ITGA2B-ceC96 and their interactions with ITB3 (*14*). First, we observed that ITGA2B-ceC96 tended to degrade in vitro (Fig. 4A). Compared with unmodified ITGA2B, ITGA2B-ceC96 was more easily degraded in cycloheximide (CHX)–treated cells (Fig. 4B). The instability of ITGA2B-ceC96 likely occurred because the Cys87–Cys96 disulfide bond was disrupted. To examine the effect

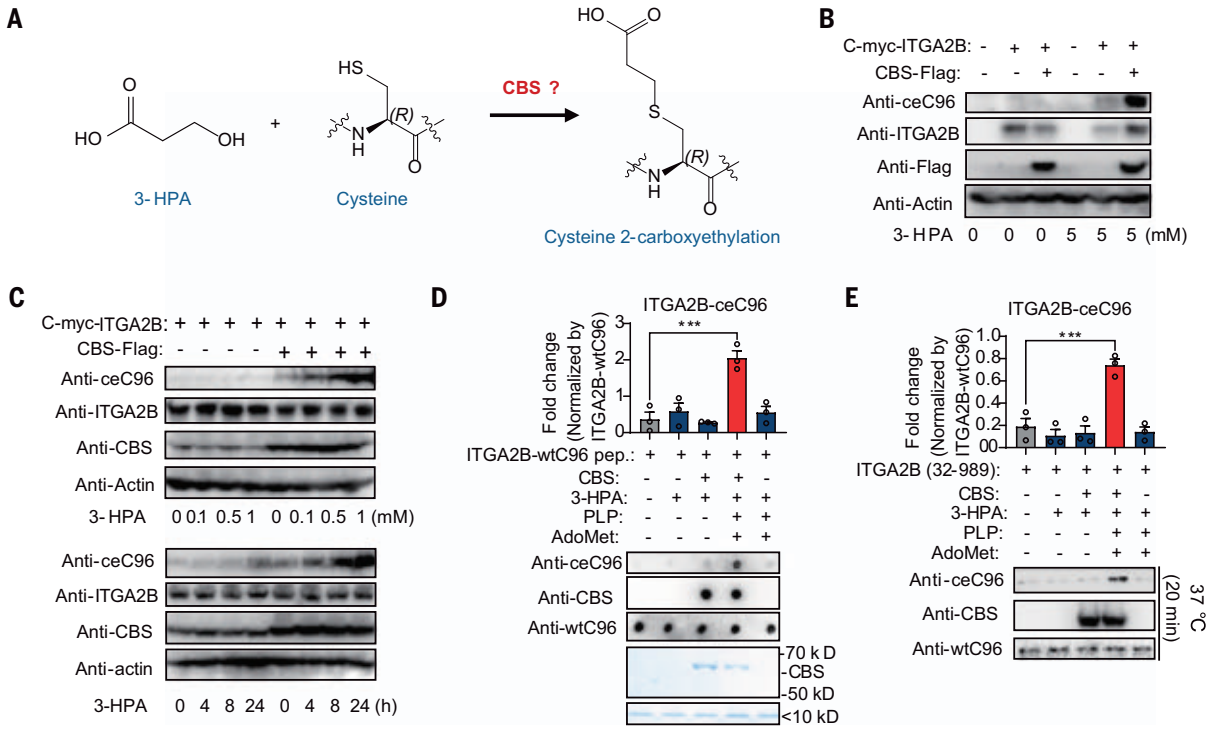


Fig. 3. Cysteine carboxyethylation is catalyzed by CBS. (**A**) For proteins that contain cysteine, CBS is expected to promote cysteine 2-carboxyethylation from cysteine and 3-HPA. (**B** and **C**) Immunoblot results obtained using anti-ceC96 antibody, anti-ITGA2B antibody, anti-FLAG antibody, and anti-CBS antibody to analyze 293T cells overexpressing FLAG-tagged CBS and C-Myc-tagged ITGA2B after treatment with 3-HPA. The results shown are representative of three independent experiments. (**D** and **E**) Reactions of unmodified ITGA2B peptide (ITGA2B-wtC96 peptide) (D) and recombinant ITGA2B protein (32–989) (E) and CBS, 3-HPA, PLP, and AdoMet in vitro at 37°C for 20 min. Data are shown as mean ± SEMs and $n = 3$ per group per experiment. These data are representative of three experiments. Statistical significance was determined using one-way ANOVA with Dunnett's multiple-comparisons test with $^{***}P < 0.001$.

of carboxyethylation on the stability of the bond, two mutations were produced to mimic the negatively charged carboxyl group (p.C96D) or to block the formation of this disulfide bond (p.C96M). These mutations enhanced the degradation of ITGA2B (fig. S4A) and reduced its interaction with ITB3 (fig. S4B). In addition, ITGA2B-ceC96 was stabilized by adding the lysosome inhibitor bafilomycin A1 (Baf.A) to the cells (Fig. 4C and fig. S4C). Furthermore, ITGA2B-ceC96 colocalized with the lysosome marker LAMP1 (fig. S4D). Thus, compared with unmodified ITGA2B, ITGA2B-ceC96 is less stable, likely because of its degradation through a lysosomal pathway.

Interactions between CBS and ITGA2B were associated with localization and cysteine carboxyethylation. ITGA2B-ceC96 intensely colocalized with nucleolin in nucleoli and with CBS in the cytoplasm (fig. S4E). ITGA2B-ceC96 was translocated to the nuclei in the presence of 3-HPA (Fig. 4D). Unmodified ITGA2B showed reduced nuclear localization (fig. S5, A to D). Previous work (*14*) has shown that ITGA2B undergoes structural changes upon activation or deactivation from the physiological resting state. Thus, modified ITGA2B may adopt different structures and form different complexes in the nucleus (fig. S5E). Moreover, the carboxyethylation of ITGA2B Cys96 appeared to cause limited effects on cell viability and proliferation, because 3-HPA treatment did not significantly affect apoptosis (Fig. 4E) and the ITGA2B mutations did not significantly alter the cell cycle distribution (Fig. 4F).

The degradation of ITGA2B-ceC96 may generate peptides that activate the HLA-restricted adaptive immune responses in patients with AS. Indeed, among the top 40 carboxyethylated proteins, the ITGA2B-ceC96 peptide exhibited some of the highest HLA-DRB1 scores (fig. S6A) (*15*). To investigate whether the ITGA2B-ceC96 peptides can function as neoantigens and trigger HLA-restricted adaptive immune responses, we developed in vitro assays to detect autoantibody production and autoreactive T cell responses. Using an enzyme-linked immunosorbent assay (ELISA)–based method, we found that the ITGA2B-ceC96–specific antibody showed relatively high avidity for the ceC96 bridge compared with the ITGA2B-wtC96–specific antibody (fig. S6B). We then used this method to assay plasma samples from 126 patients with AS and 40 healthy donors. Compared with the healthy donor group, the AS group showed significantly higher levels of autoantibodies against ITGA2B-ceC96. Furthermore, at least 10 patients in the AS group exhibited levels above the highest level in the healthy donor group (Fig. 5A). All 10 of these patients carried the HLA-DRB1*04 allele (Fig. 5B), which is known to be associated with AS (*16*). $CD4^+$ T cell responses were significantly induced by ITGA2B-ceC96(84-110)–pulsed mature dendritic cells (mDCs) (Fig. 5, C and D, and figs. S6C and S7, A and B) in autoantibody-positive patients. When pulsed with ITGA2B-ceC96(84-110), HLA-DRB1*04^+ mDCs showed higher HLA-DR membrane expression than mDCs carrying other HLA alleles (fig. S7C). In addition, ITGA2B-ceC96(84-110)–pulsed mDCs promoted $CD4^+$ T cell proliferation to a greater extent than ITGA2B-wtC96(84-110)–pulsed mDCs (fig. S7D).

$CD4^+$ T cells stimulated with ITGA2B-ceC96(84-110)–pulsed mDCs were then isolated by fluorescence-activated cell sorting (FACS) (fig. S8, A and B) and their T cell receptor (TCR) repertoires were determined by single-cell TCR sequencing (fig. S8C and data S2). Four TCR clonotypes were validated by retroviral expression in a $CD4^+$ Jurkat cell line depleted endogenous TCRs (Jurkat-TCR^- cells) (*17*, *18*). Cell surface expression of CD3 was used to validate the successful expression of exogenous TCRs, and an HLA-DRB1*04–ceC96 tetramer was used to verify TCR and peptide–major histocompatibility complex (MHC) interactions (fig. S8D). Using this approach, we verified that TCR clone #4 could bind the HLA-DRB1*04–ceC96 tetramer (fig. S8E). To verify this TCR's

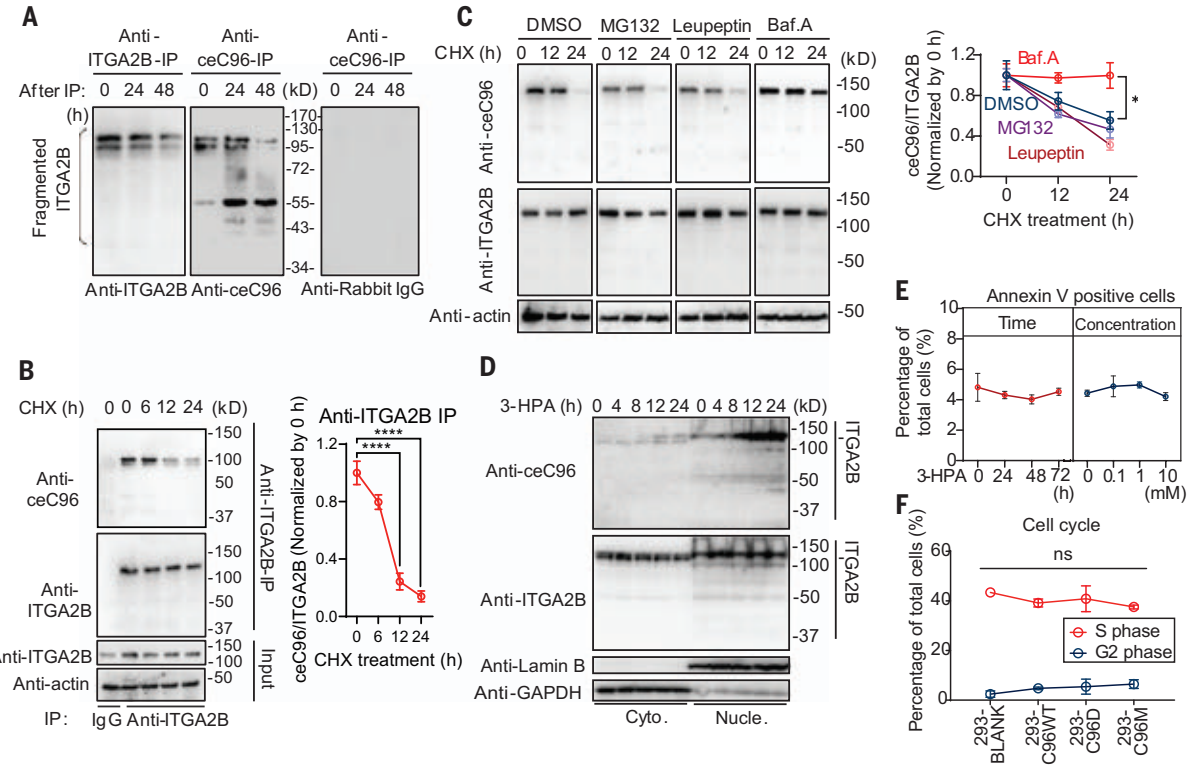


Fig. 4. ITGA2B cysteine carboxyethylation induces lysosomal degradation. (**A**) Stability of carboxyethylated ITGA2B and unmodified ITGA2B after immunoprecipitation from patient PBMCs. Full immunoblots after staining an anti-ITGA2B antibody (total ITGA2B protein), anti-ceC96 antibody (carboxyethylated ITGA2B), and anti-rabbit IgG antibody (control for nonspecific antibody reactivity during immunoprecipitation) at 0, 24, or 48 hours after immunoprecipitation. (**B**) Immunoblot of ITGA2B immunoprecipitation after CHX treatment. The anti-ceC96 antibody and anti-total ITGA2B antibody were used to compare the degradation rates of carboxyethylated ITGA2B and total ITGA2B. The first lane shows mouse IgG immunoprecipitation as a negative control. $n = 3$ per group per experiment. These data are representative of two experiments. (**C**) Immunoblot of cysteine carboxyethylated ITGA2B degradation after treatment with dimethyl sulfoxide, MG132, leupeptin, or bafilomycin A1 combined with CHX ($n = 3$). Cells were pretreated with 3-HPA for 24 hours, and then the medium was replaced before the CHX, MG132, leupeptin, and Baf.A treatments. (**D**) Immunoblot depicting the nuclear and cytoplasmic localization of ITGA2B with carboxyethylated cysteine in the presence of the modifier 3-HPA. The results shown are representative of three independent experiments. (**E**) Apoptosis rate measured by annexin V staining after treatment with 3-HPA at different concentrations for 0, 24, 48, or 72 hours ($n = 3$). (**F**) Cell cycle distribution of 293T cells transfected with wtC96 (C96WT) or mutant wtC96 (C96D or C96M) ($n = 3$). Data are shown as mean ± SEMs and are representative of three experiments. Statistical significance was determined using two-way ANOVA with Tukey's multiple-comparisons test. ns, Not significant.

antigen specificity, mDCs were isolated from 12 AS patients. Patients DC1 to DC6 were HLA-DRB1*04^{+}, whereas patients DC7 to DC12 were HLA-DRB1*04^{-}. DCs from all clinical samples were pulsed with either ITGA2B-ceC96(84-110) (samples 1 to 12) or ITGA2B-wtC96(84-110) peptides (samples 16 to 27) (table S4). TCR #4 prompted interleukin-2 (IL-2) production by Jurkat cells only upon stimulation by HLA-DRB1*04^{+} mDCs pulsed with the ITGA2B-ceC96(84-110) peptide (fig. S8F). Thus, the ITGA2B-ceC96(84-110) peptide promotes an antigen-specific T cell response that is restricted to HLA-DRB1*04, possibly because this HLA favors the presentation of the modified peptide.

*The ITGA2B-ceC96 peptide is specifically recruited to the antigen-binding groove of the HLA-DRA*01–HLA-DRB1*04 complex*

To confirm the presentation of the ITGA2B-ceC96 peptides by HLA-DRB1*04 and the downstream response, an in vitro presentation system containing HLA-DR and HLA-DM complexes was constructed and verified (fig. S9). The ITGA2B-ceC96(84-110) peptide, named ceC96-27, was recruited to the HLA-DRB1*04/HLA-DRA*01 complex in the presence of HLA-DM. Additional evidence of this interaction was provided by LC–MS/MS analysis (fig. S10, A to C).

To identify HLA-DR–presented peptides, the fluorescence polarization (FP) of peptides containing ITGA2B Cys96 was analyzed and compared with that of a positive control (pep-PC: PKYVKQNTLKLAT) (*19*) (Fig. 6A). ITGA2B-ceC96(85-98) (ceC96-F, FLCPWRAEGGQC$_{ce}$PS) showed the largest difference in FP, indicating the strongest binding to the HLA-DRB1*04–HLA-DRA*01 complex (Fig. 6B). This fragment was then aligned with the antigen-binding pocket of the HLA complex using UCSF Chimera–based modeling (*20*). A three-dimensional model for HLA-DRA*01/HLA-DRB1*04:02:01 and the ITGA2B-ceC96(85-98) (wtC96-F) (fig. S10D) was built over the template of a previously reported structure of HLA-DRB1*04:01/α-enolase 26/cit26 peptides (*21*). After minimization, the final structure showed highly conserved binding pockets for the wild-type epitope (fig. S10, E to G). According to this model, the same peptide with a carboxyethylated Cys96 could form new hydrogen bonds with Trp^{60} and Gln^{64} (fig. S10H, top) in the HLA molecule beta chain or with Gln^{95} (fig. S10H, middle) within the same peptide after minimization. This peptide could also dock with its carboxyethyl group side chain at Cys^{96} pointing out of the HLA-binding pocket (fig. S10H, bottom). Thus, Cys^{96} carboxyethylation can potentially influence binding to HLA class II molecules and alter the molecular surface available for TCR recognition.

We next explored whether this peptide could be presented as a neoantigen. The ceC96 peptide colocalized with HLA-DR on the surface of patient-derived DCs (Fig. 6C), but no surface colocalization was observed with HLA-A/B/C (Fig. 6D). Thus, carboxyethylation of ITGA2B

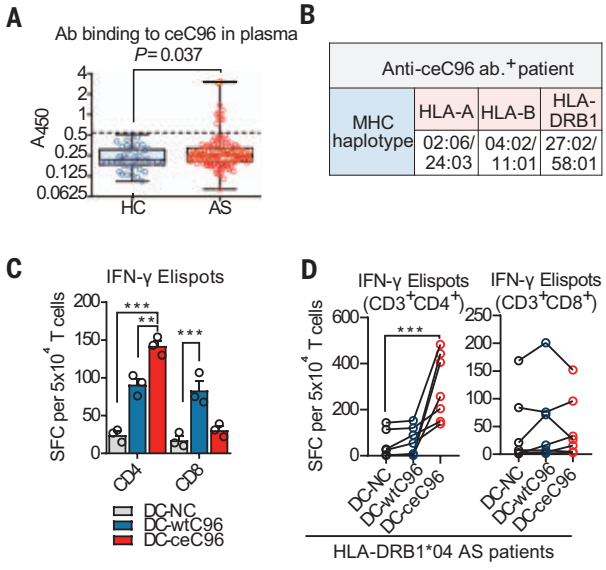


Fig. 5. ITGA2B cysteine carboxyethylation induces CD4$^+$ T cell-related immune responses. (**A**) Anti-ceC96 antibodies in the plasma of AS patients (n = 126) and HCs (n = 40) were quantified by ELISA. The cutoff value (dotted line) is three SDs above the mean of the HC group. Data are shown as mean ± SDs and are representative of two experiments. Statistical significance was determined using Student's t test. (**B**) MHC typing (HLA-A/B/DRB1) of the patient indicated by a yellow circle in (A) with high levels of anti-ITGA2B-ceC96 antibody. (**C**) IFN-γ$^+$ cells from the patients were measured by ELISpot assay. n = 3 per group per experiment. These data are representative of two experiments. DC-NC (negative control), DC-wtC96, and DC-ceC96 indicate T cells stimulated with original DCs, DCs pulsed with the wtC96 peptide, and DCs pulsed with the ceC96 peptide, respectively. SFC, Spot-forming cells. (**D**) IFN-γ ELISpot assay for cells from AS patients (n = 7) who were HLA-DRB1*04$^+$. For (C) and (D), data are shown as mean ± SEMs and are representative of three experiments. Statistical significance was determined using one-way ANOVA with Dunnett's multiple-comparisons test with $^{**}P < 0.01$ and $^{***}P < 0.001$.

at Cys^{96} leads to the production of neoantigens that are presented by the HLA-DRB1*04/HLA-DRA*01 complex and bind to HLA-DR on the surface of patient-derived DCs.

HLA-DR4–transgenic mice, which feature the same antigen-binding specificity as HLA-DRB1*04:01–positive individuals (*22*), were immunized with ITGA2B-ceC96(84-110) and ITGA2B-wtC96(84-110) antigenic peptides (fig. S11A). On days 21 and 31 after immunization, antibodies against ITGA2B-ceC96 were present in only the mice immunized with the ITGA2B-ceC96(84-110) peptide, not in those immunized with ITGA2B-wtC96(84-110) (Fig. 6E and fig. S11B). Moreover, the percentage of CD138$^+$ plasma cells and effector CD4$^+$ T cells increased in the mice immunized with ITGA2B-ceC96(84-110) (Fig. 6, F and G), and did not change in ITGA2B-wtC96(84-110)–immunized mice (fig. S11, C to E).

*ITGA2B cysteine carboxyethylation is associated with autoantibodies and T cell responses in HLA-DRB1*04 patients with AS*

We further examined autoantibodies specific for cysteine carboxyethylation in HLA-DRB1*04$^+$ patients. A total of 103 AS patients were typed for HLA-A, HLA-B, and HLA-DRB1 (Fig. 7, A to C). Consistent with previous studies (*16*), the frequency of HLA-DRB1*04 was 33.01% among patients with AS (Fig. 7C) and <23% among healthy populations (Fig. 7D). Mean anti–ITGA2B-ceC96 antibody levels were higher in HLA-DRB1*04$^+$ patients than in HLA-DRB1*04$^-$ patients (Fig. 7E). By contrast, there was no correlation between HLA-B*27 and HLA-A*02 alleles and anti–ITGA2B-ceC96 antibody production (Fig. 7, F and G). HLA-DRB1*04$^+$ patient–derived DCs, which were pulsed with ITGA2B-ceC96 peptide (DC-ceC96), significantly induced CD4$^+$ T cell responses, as indicated by the increased percentages of CD4$^+$ T cells expressing membrane CD137 and intracellular interferon γ (IFN-γ), tumor necrosis factor α (TNF-α), and IL-2 (Fig. 7H), compared with DCs pulsed with no peptide (DC-NC) or the ITGA2B-wtC96 peptide (DC-wtC96). By contrast, DC-ceC96 from HLA-DRB1*04$^-$ patients did not induce CD4$^+$ T cell responses. Furthermore, there were no significant differences in DC-induced CD8$^+$ T cell responses among all three groups from HLA-DRB1*04$^-$ or HLA-DRB1*04$^+$ patients (fig. S11F). Thus, ITGA2B cysteine carboxyethylation is associated with autoantibody production and antigen-specific CD4$^+$ T cell responses in HLA-DRB1*04$^+$ patients.

ITGA2B-ceC96–related metabolic and immune changes are present in AS, rheumatoid arthritis, and systemic lupus erythematous patients as well as healthy donors

To further investigate the origin of cysteine carboxyethylation, we analyzed a total of 110 PBMC samples from patients with AS, rheumatoid arthritis (RA), and systemic lupus erythematous (SLE), as well as from healthy controls (HCs). We also sought to ascertain the prevalence of and possible correlations among ITGA2B-ceC96 antigens, autoreactivity to ITGA2B-ceC96, HLA-DRB1*04:02 expression, and autoimmune disease.

Carboxyethylated ITGA2B was detected in PBMCs from patients with AS, RA, and SLE. Specifically, carboxyethylated ITGA2B levels in PBMCs were significantly higher in patients with AS than in healthy donors (Fig. 8A and fig. S12, A to D). However, carboxyethylated ITGA2B was not detected in synovial fluid cells of patients with AS (fig. S12E). Possible correlations among autoreactivity to ITGA2B-ceC96, HLA-DRB1*04:02 expression, and autoimmune diseases were then examined. HLA-DRB1*04-ceC96 2D tetramer$^+$ cells (indicating autoreactive T cells) patients were identified among the AS, RA, and SLE groups (Fig. 8B and fig. S13A), and most tetramer$^+$ patients were also HLA-DRB1*04$^+$ (fig. S13B). Additionally, ITGA2B-ceC96–specific autoantibody levels were significantly increased in the plasma of patients with AS, RA, and SLE compared with that of the HCs (Fig. 8C), but ITGA2B-wtC96–specific autoantibody levels were comparable among the groups. Neither sex nor age correlated with the level of ITGA2B-ceC96–specific autoantibodies (fig. S14, A to C).

3-HPA levels were significantly increased in plasma samples from patients with AS compared with those from the HCs and were relatively higher in anti-ceC96 autoantibody–positive patients (Fig. 8D and fig. S15, A and B). Among the metabolites involved in the 3-HPA–related pathway, only propionic acid and citric acid were lower in patients with AS (fig. S15B). CBS levels were increased in patients with AS but not in those with RA or SLE (Fig. 8E and fig. S12, A to D). Thus, the 3-HPA and related metabolic pathways are enhanced significantly in patients with AS. Furthermore, positive correlations trended in patients with RA and SLE compared with HCs.

The data for AS, RA, and SLE patients are quantitated in table S5. ITGA2B-ceC96 levels (fig. S16A) correlated with levels of 3-HPA and anti-ceC96 autoantibody in the plasma, especially in the HLA-DRB1*04$^+$ patients (fig. S16, B and C), suggesting that the HLA-DRB1 haplotype is a possible risk factor for AS pathogenesis. Thus, 3-HPA, a key driver of cysteine carboxyethylation, and ITGA2B-ceC96–specific immune responses appear to play important roles in AS.

Although AS primarily affects the spine and sacroiliac joints, which are sites difficult to biopsy in humans because of ethical issues, one of the most common extraskeletal manifestations is inflammatory bowel disease (*9*, *23*). Indeed, we detected ITGA2B-ceC96 in colon tissue sections from AS patients (fig. S17, A and B). Immunized HLA-DR4 mice lost weight and developed colitis at 8 weeks after the first immunization with the wtC96 or ceC96 peptide conjugated to the carrier protein keyhole limpet hemocyanin (KLH) (fig. S18, A and B). Colon histological scores and disease activity index (*24*) were higher in KLH-ceC96–immunized and 3-HPA–treated mice (fig. S18, C to F) than in those immunized with the wtC96 peptides. The fecal levels of lipocalin-2 (Lcn2)

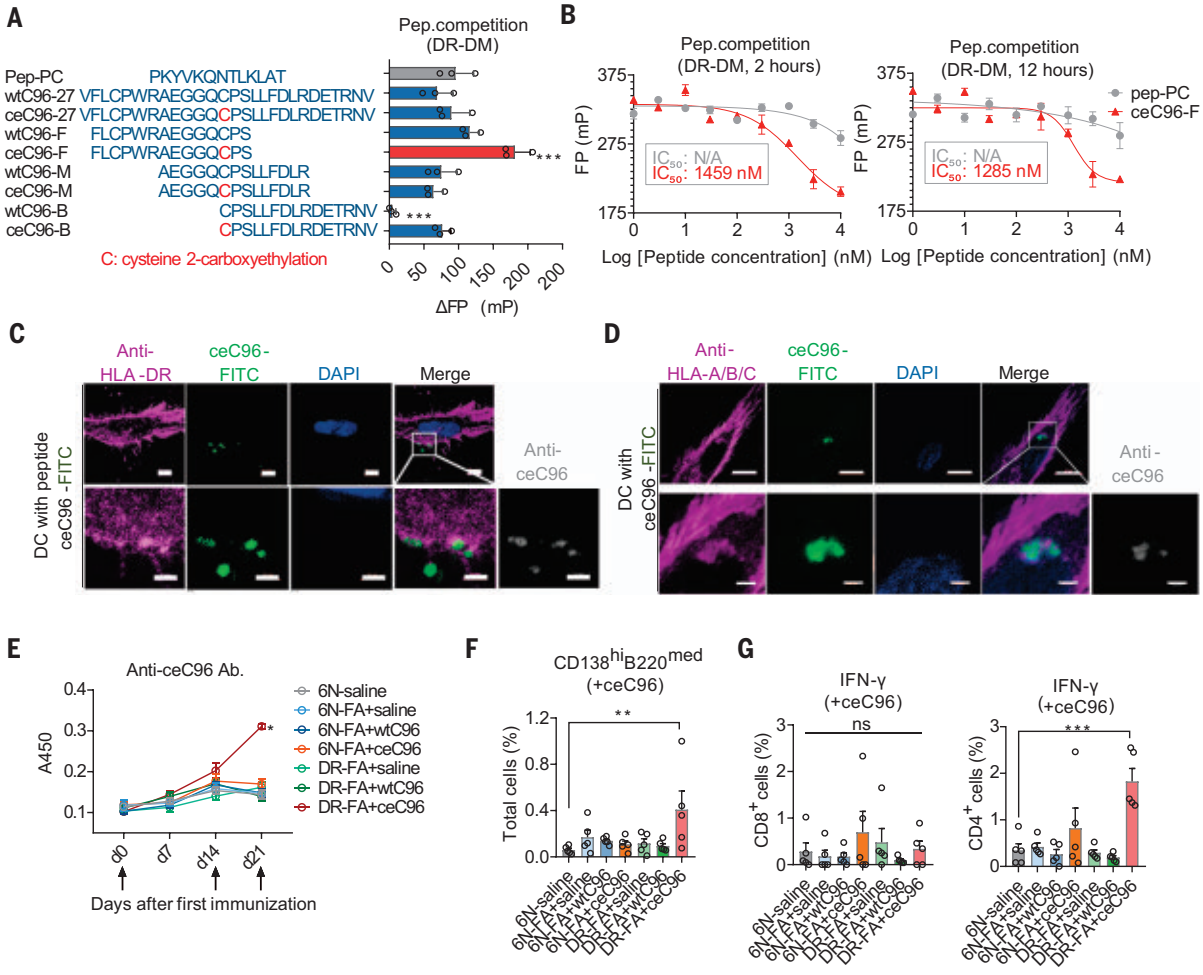


Fig. 6. HLA-DR4–restricted immune responses are triggered by carboxyethylated ITGA2B peptides. (**A**) Candidate peptide sequences and FP results for peptide competition assays. An HLA-DR/FAM-prepeptide complex was generated, and unbound peptides were removed with a gel-filtration column. All competition peptides (1 μM) were added to the HLA-DR/FAM-prepeptide complex (100 nM) with HLA-DM (30 nM) for 2 hours. Influenza A HA (307–319) was used as the positive control peptide (pep-PC). ΔFP values were obtained by subtracting the competition group FP values from the HLA-DR/FAM-prepeptide complex–only group FP values. Data are shown as mean ± SEMs and are representative of three experiments. Statistical significance was determined using one-way ANOVA with Dunnett's multiple-comparisons test with $^{***}P < 0.001$ compared with the pep-PC group. (**B**) FP results for peptide competition (pep-PC and ceC96-F) performed for 2 hours and 12 hours at different peptide concentrations ($n = 3$). Data are presented as the mean ± SEMs and are representative of two experiments. Statistical significance was determined using four-parameter curve fittings, and the median inhibitory concentration (IC_{50}) for ceC96-F after 2 hours was 1459 nM, whereas the IC_{50} for ceC96-F after 12 hours was 1285 nM. (**C** and **D**) Representative images of confocal immunofluorescence analysis ($n = 5$) of the colocalization of the modified peptide (ceC96-FITC) (green) and HLA-DR/HLA-A/B/C (magenta) on the cell membrane. An anti-ceC96 antibody was used to verify the presence of the ceC96 peptide (gray). Scale bars: top panel, 5 μm; bottom panel, 2 μm. $n = 5$ per group per experiment. These data are representative of three experiments. (**E**) Detection of the anti–ITGA2B-ceC96 antibody in mouse plasma ($n = 5$ for each group) on days 0, 7, 14, and 21. The arrows indicate the days on which the three immunizations were performed. Data are shown as mean ± SEMs and are representative of three experiments. Statistical significance was determined using two-way ANOVA with Tukey's multiple-comparisons test with $^{*}P < 0.05$. (**F**) Changes in the percentage of $CD138^{+}$ plasma cells after coculture with ceC96 peptides. Data are shown as mean ± SEMs and are representative of three experiments. Statistical significance ($n = 5$) was determined using one-way ANOVA with Dunnett's multiple-comparisons test with $^{**}P < 0.01$. (**G**) The percentage of IFN-γ^{+} T cells after coculture with ceC96 peptides ($n = 5$). Data are shown as mean ± SEMs and are representative of three experiments. Statistical significance ($n = 5$ for each group) was determined using one-way ANOVA with Dunnett's multiple-comparisons test with $^{***}P < 0.001$. ns, not significant.

and the percentage of animals positive for fecal occult blood were both extremely elevated in KLH-ceC96–immunized and 3-HPA–treated mice (fig. S18, G and H). Moreover, $CD11c^{+}$ cells (mainly macrophages and DCs and some granulocytes and B cells), $Ly6G^{+}$ cells (mainly neutrophils and some monocytes and other granulocytes), and $CD4^{+}$ T cells were detected in the colonic lamina propria, which partly reflected the increased level of inflammation in the colons of KLH-ceC96–immunized and 3-HPA–treated mice (fig. S19). Finally, ITGA2B-ceC96–positive cells, ITGA2B-ceC96–specific T cells and total immunoglobulin G (IgG) were detected in the colons of KLH-ceC96–immunized and 3-HPA–treated mice (figs. S20 and S21). Thus, ITGA2B-ceC96–specific $CD4^{+}$ T cell responses appeared to be associated with the pathogenesis of colitis in this mouse model.

Significant bone erosion and remodeling were detected throughout the vertebrae of ITGA2B-ceC96(84-110)–immunized and 3-HPA–treated mice, particularly in the lumbar spine (fig. S22, A to D). Hematoxylin & eosin (H&E) and safranin O-fast green staining indicated the presence of bone erosion (trabecular bone

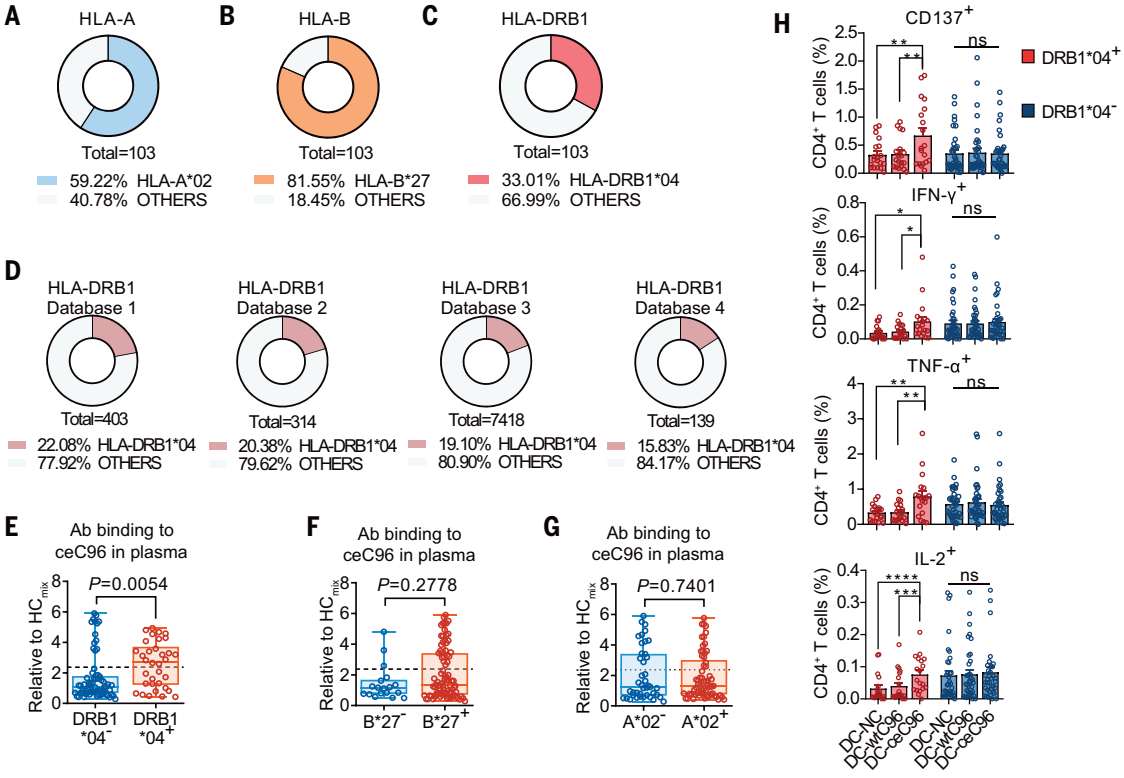


Fig. 7. ITGA2B-ceC96 peptide-related T cell response and autoantibody production in AS patients with specific HLA haplotypes. (**A** to **C**) Distribution of the HLA-A*02 (A), HLA-B*27 (B), and HLA-DRB1*04 (C) alleles in 103 patients. (**D**) The percentage of HLA-DRB1*04^{+} individuals from the Allele*Frequencies database (*57*–*59*) (www.allelefrequencies.net, pop_id = 2410, 2007, 2668, and 1820, respectively). (**E**) Anti-ITGA2B-ceC96 antibody production in HLA-DRB1*04^{+} (n = 34) and HLA-DRB1*04^{-} (n = 69) patients as determined by ELISA. (**F**) Anti-ITGA2B-ceC96 antibody production in HLA-B*27^{+} (n = 85) and HLA-B*27^{-} (n = 18) patients as determined by ELISA. (**G**) Anti-ITGA2B-ceC96 antibody production in HLA-A*02^{+} (n = 61) and HLA-A*02^{-} (n = 42) patients as determined by ELISA. In (E) to (G), data are shown as mean ± SDs and are representative of two experiments. Statistical significance was determined using Student's t test. The cutoff value (dotted line) represents a threefold increase above the mean in the control group value. (**H**) The percentage of CD137^{+}, IFN-γ^{+}, TNF-α^{+}, and IL-2^{+} CD4^{+} T cells was determined by flow cytometry in HLA-DRB1*04^{+} (n = 19) and HLA-DRB1*04^{-} (n = 37) patients. Data are shown as mean ± SEMs and are representative of two experiments. One-way ANOVA with Dunnett's multiple-comparisons test with *P < 0.05; **P < 0.01; ***P < 0.001; and ****P < 0.0001; ns, not significant.

loss) and enthesitis in diseased bones (fig. S22E). Thus, ITGA2B-ceC96 immunization leads to increased susceptibility to spondylitis, a skeletal manifestation of AS.

Discussion

We systematically screened possible PTMs across the whole proteome and identified cysteine carboxyethylation as a specific protein modification in the PBMCs of patients with AS. Cysteine carboxyethylation promoted lysosomal degradation of ITGA2B, generating HLA-DRB1*04–restricted neoantigens. These carboxyethylated neoantigens appeared to be pathogenic and induced colitis in HLA-DR4 mice, a typical extra-articular inflammatory manifestation of AS. Metabolic disorders that are characterized by increased 3-HPA promoted cysteine carboxyethylation. We propose that cysteine carboxyethylation has a syngeneic effect on metabolic disorders and aberrant immune tolerance that may drive autoimmune diseases such as AS.

The pathogenesis of AS is not well understood. Autoantibodies are not often detected in patients with AS, and whether AS is a T cell–driven autoimmune disease remains controversial (*9*). In our mouse model, carboxyethylated antigens were necessary for the development of colitis and spondylitis. Moreover, 3-HPA, the substrate for cysteine carboxyethylation, participates in several metabolic pathways that are associated with autoimmune diseases (*25*–*27*), and plasma 3-HPA levels have been found to be increased in patients with AS. Consistently, our data demonstrate that cysteine carboxyethylation is an enzymatic reaction that requires 3-HPA in vitro and in vivo. Exogenous 3-HPA is derived mainly from microbes such as *Escherichia coli* (*28*) and *Klebsiella pneumonia* (*29*) and from food products containing Poaceae, Fabaceae, and *Saccharomyces cerevisiae* (*30*). These observations are also consistent with evidence that the ITGA2B-ceC96 antigen is present in the gut of patients with AS, which is known to be related to intestinal disorders: 5 to 10% of AS patients are clinically diagnosed with inflammatory bowel disease, and 70% are clinically diagnosed with subclinical gut inflammation (*8*, *9*). Thus, our study agrees with a general pathogenic pathway in which metabolic changes could induce protein modifications (*31*–*34*) and produce pathogenic neoantigens, leading to autoimmunity.

Cysteine is the most active residue among the standard amino acids, but only a small number of cysteine modifications have been reported because of limitations in the methodology in modern protein technologies. Our work provides a workflow to profile cysteine modifications and disease-associated amino acid derivatives that appear to be important for neoantigen production in different immune disorders. This workflow provides a general approach to systematically screen PTM-related neoantigens in patients with autoimmune diseases.

Materials and methods

Clinical patient samples and cell lines

All clinical samples were obtained from Xijing Hospital, Fourth Military Medical University (Xi'an, China). Information about clinical samples is provided in table S1. All patients were diagnosed according to the 1984 modified New York criteria for AS (*35*), the American Rheumatism Association 1987 revised criteria for the classification of RA (*36*), and the American College of Rheumatology 1997 revised classification criteria for SLE (*37*). Informed written consent was obtained from all patients or their families and from all healthy individuals before participation, and ethics approval for this study was granted by the Ethics Committee of Fourth Military Medical University (no. 20110303-7). THP-1 and human embryonic kidney 293T (HEK 293T) cell lines (table S6) purchased from ATCC were

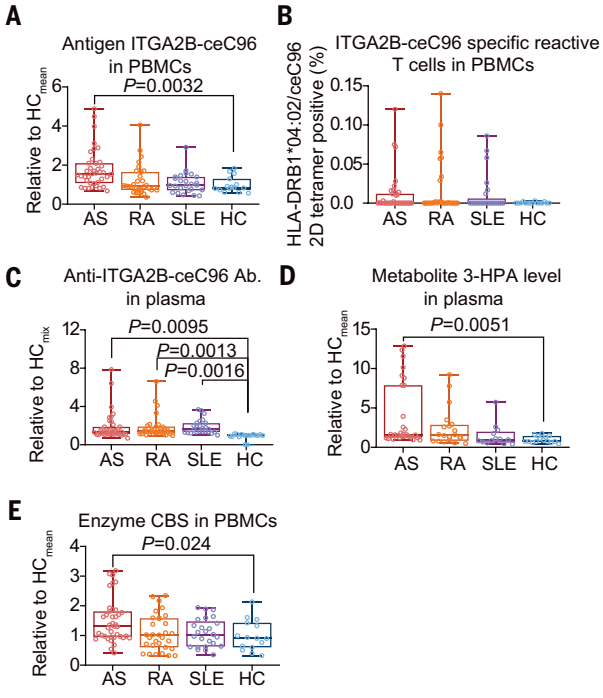


Fig. 8. Comprehensive analysis of ITGA2B-ceC96–related metabolic and immune changes in AS, RA, and SLE patients and HCs. (**A**) The levels of the antigen ceC96 in PBMCs of AS (n = 39), RA (n = 31), and SLE (n = 25) patients and HCs (n = 15) quantified by immunoblotting assays. (**B**) The ITGA2B-ceC96–specific T cell levels in PBMCs of AS (n = 39), RA (n = 31), and SLE (n = 25) patients and HCs (n = 15) were quantified by flow cytometry analysis. (**C**) Anti–ITGA2B-ceC96 antibody titers in plasma of AS (n = 39), RA (n = 31), and SLE (n = 25) patients and HCs (n = 15) were quantified by ELISA. (**D**) 3-HPA levels in plasma of AS (n = 29), RA (n = 20), and SLE (n = 13) patients and HCs (n = 15) were quantified by mass spectrometry. (**E**) CBS levels in PBMCs of AS (n = 39), RA (n = 31), and SLE (n = 25) patients and HCs (n = 15) quantified by immunoblot assays. In (A) to (E), ITGA2B-ceC96, 3-HPA, and CBS levels were normalized to the corresponding mean value of the HCs (HC_{mean}) and anti–ITGA2B-ceC96 antibody levels were normalized to the value of HC_{mix}. Data are shown as mean ± SDs and are representative of two experiments. Statistical significance was determined using one-way ANOVA with Dunnett's multiple-comparisons test, with the P value displayed in the figure.

cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS), 1% penicillin–streptomycin, and 2% L-glutamine at 37°C in a humidified atmosphere with 5% CO_2.

LC–MS/MS sample preparation

Human PBMC samples were obtained from seven AS patients and five HCs. Total protein was then isolated in the presence of a protease inhibitor cocktail (Roche, Switzerland) and a phosphatase inhibitor cocktail (Roche, Switzerland). The cell lysates were centrifuged at 10,000g, the supernatants were collected, and protein levels were measured with a BCA kit (Pierce, Thermo Scientific, Germany). To preserve the biochemical properties of native protein residues during protein preparation, reagents such as iodoacetamide and urea, which potentially modify proteins, were omitted. Two hundred micrograms of cell lysate was digested with trypsin. The samples were dialyzed with NH_4HCO_3 and reduced with DL-dithiothreitol (Sigma-Aldrich, USA). ZipTip C18 spin columns (Millipore, USA) were used to purify the peptides (*38*).

Mass spectrometry and protein sequence alignment analyses

Mixed peptides were fractionated on a reverse-phase C18 column by high-performance liquid chromatography as previously described (*38*–*40*). The desalted peptides were analyzed by LC–MS/MS on an LTQ-Orbitrap Elite mass spectrometer (Thermo Scientific) in data-dependent mode, in which MS/MS fragmentation of the 20 most intense peaks was acquired for every full MS scan. MS/MS spectra were searched against the human protein database UniProt using SEQUEST (*41*), with a maximum allowed deviation of 10 ppm for the precursor mass and 0.6 Da for fragment masses. Dynamic modification included the oxidation of methionine, and the false discovery rate (FDR) was 1%. Every sample was run in triplicate. A multiblind spectral alignment algorithm, Byonic (*42*, *43*), was used for open modification searching. The modification was set from −200 to 1000 Da, and the fragment tolerance was 0.6 Da. For clustering, mass shifts (delta masses) were divided into subgroups with 1-Da intervals bounded by n ± 0.5 (n = −200 to 1000) Da. The mass shifts in each mass window were analyzed by multivariate clustering using a Gaussian mixture model. The clusters within each window were determined by the Bayesian information criterion (*44*). Next, the clusters in each window were fitted individually by Gaussian regression to calculate the peak value (expected mass shift), the SD, and the goodness of fit (R^2).

Modified substrate screening

The unmodified peptide wtC96 (30 mM) (AEGGQCPSLLFDLR) was incubated with either LA or 3-HPA (1 or 5 mM) in 20 μg of 293T cell lysate at 37°C for 2 hours. LC–MS/MS was performed to identify the modified peptide (C96+72.021).

Stability assay with modified peptides and mutant ITGA2B

To test the effect of pH on stability, 0.5 mg of modified peptide (#5 or #7) was incubated at 37°C for 16 hours at pH 5.0, 6.0, 6.5, or 7.4. After incubation, all samples were identified by LC–MS/MS. To test the effect of ITGA2B mutation on stability, the cysteine (C) at residue 96 in ITGA2B was mutated to aspartic acid (D) or methionine (M). ITGA2B wtC96 (C96WT), C96D, and C96M cDNAs were cloned into the expression vector pCMV3, and the resulting vectors were transfected into the 293T cells. After 48 hours, CHX (2 μg/ml) was added to 293T cells for 0, 6, or 12 hours. 293T cells were treated with MG132 (5 μM), leupeptin (100 μM), and bafilomycin A1 (5 nM) to block protein degradation through different pathways.

Anti-ITGA2B-ceC96–specific antibody generation and verification

Anti-ITGA2B-ceC96 (anti-ceC96) IgG was developed in our laboratory as follows. ITGA2B-ceC96 peptide was coupled with KLH at a 3:1 ratio (AEGGQC_{ce}PSLLFDLRC-KLH). A total of 0.75 mg of KLH-conjugated ITGA2B-ceC96 peptide dissolved in double-distilled H_2O and emulsified (1:1) in complete Fraud's adjuvant (CFA) was used to immunize rabbits once, and the rabbits were then boosted three times with the peptide in incomplete Fraud's adjuvant (IFA) at 2-week intervals. Five batches of serum were collected from each rabbit. The serum sample with the highest ELISA titer was used to enrich the anti-ITGA2B-ceC96 antibody. The modified peptide (ITGA2B-ceC96-BSA, AEGGQC_{ce}PSLLFDLRC-BSA) and unmodified peptide (ITGA2B wtC96-BSA, AEGGQCPSLLFDLRC-BSA) were diluted with 0.1 M $NaHCO_3$ and 0.5 M NaCl buffer (pH 8.3) to a concentration of 5 mg/ml. Then, 5 mg of protein was added to 10 ml of agarose solution. The samples were shaken overnight at 4°C, washed three times with 0.1 M $NaHCO_3$ and 0.5 M NaCl buffer, added to a 3× volume of blocking buffer (0.1 M Tris-HCl, pH 8.0), shaken at 4°C for 2 hours, and loaded onto the columns. Two hundred milliliters of rabbit serum ammonium sulfate precipitation mixture was centrifuged at 5000g for 30 min, the supernatant was discarded, and the precipitated protein was dissolved in 0.01 M phosphate-buffered saline (PBS) buffer (pH 7.4) and filtered through a 0.22-μm filter. Gel-filtration chromatography was used to repeat the ITGA2B-wtC96 prepurification process. The protein solution was collected until the absorbance was <10 mAU. A gel-filtration column for ITGA2B-ceC96 purification was equilibrated 10 times with equilibration buffer (0.01 M PBS, pH 7.4). After

loading, the column was washed to the baseline, eluted with 0.05 M Gly-HCl (pH 2.5), and neutralized (to pH 7.0) with 3 M Tris-HCl buffer. The antibody was concentrated with a 10-kD ultrafiltration tube and stored in 0.01 M PBS (pH 7.4).

Antibody specificity was verified with dot blot and surface plasmon resonance (SPR) assays. For the dot blot assay, 5, 25, and 100 ng of peptide was spotted on a nitrocellulose membrane, which was then blocked in 5% bovine serum albumin (BSA) in Tris-buffered saline and Tween 20 (TBS-T) for 1 hour at room temperature (RT) and incubated with primary antibody (anti–ITGA2B-ceC96 antibody, 0.1 μg/ml) in BSA/TBS-T for 2 hours at RT. The peptides were visualized with goat anti-rabbit IgG–horseradish peroxidase (HRP) secondary antibody (1:10,000). For the competition assay, ITGA2B-ceC96 peptide at a final concentration of 0.1 mg/ml was added to the primary antibody before incubation. Antibody affinity was tested with an SPR assay (*45*). The anti–ITGA2B-ceC96 antibody was immobilized on the chips, and wtC96 and ceC96 peptides were injected onto the chip for verification.

Screening and verification of the modifying enzymes

For screening, an anti-ITGA2B antibody (0.1 mg/ml, Santa Cruz Biotechnology) and anti–ITGA2B-ceC96 antibody (0.1 mg/ml) were used to coimmunoprecipitate a protein complex from the lysates of 293T cells overexpressing ITGA2B. The differential components in the two protein complexes were identified by LC–MS/MS, and the potential modification enzymes were predicted by protein protein interaction networks using STRING. For verification, the cDNA sequence of CBS-FL (1 to 551 amino acids) was subcloned into the pET21a(+) vector with a C-terminal His tag. ITGA2B cDNA (32 to 488 or 32 to 989 amino acids) was subcloned into the expression vector pET21b. The proteins were overexpressed in *E. coli* Rosetta (DE3) cells and purified by Ni-NTA affinity chromatography (QIAGEN). CBS protein needed to be purified in the presence of 0.1 mM PLP. The recombinant ITGA2B and CBS proteins were examined by an in vitro enzymatic reaction. The 25-μl reaction mixture contained 250 mM Tris (pH 8.6), 10 mM 3-HPA, 0.5 mM AdoMet, 0.25 mM PLP, and 30 μg of recombinant ITGA2B protein or synthetic peptide. After preincubation at 37°C for 5 min, 0.5 μg of recombinant CBS protein was added to start the reaction, which was terminated after 20 min or 2 hours by adding 25 μl of 10% trifluoroacetic acid. The presence of cysteine 2-carboxyethylation was verified by dot blot assays and immunoblotting assays.

Immunofluorescence and multicolor immunohistochemistry imaging

For immunofluorescence, cells were washed twice with PBS and stained as previously described (*46*). The concentration of the antibody used in this study and the dilution ratios of commercial antibodies are listed in table S6. The cells were visualized using an IX73 microscope system with a dry 20/0.75 lens (Olympus, Tokyo, Japan) and an A1R-A1 confocal laser microscope system (Nikon, Tokyo, Japan) with an oil 100/1.49 lens. Multicolor-IHC was performed using the Opal 7-Color Manual IHC Kit according to the manufacturer's protocol (Akoya Biosciences). The concentrations of the antibodies are listed in table S6. Multicolor Opal slides were visualized using the PhenoImager HT (Akoya Biosciences) using a sCMOS camera with a dry 20/0.75 lens and Vectra Quantitative Pathology Imaging Systems (Akoya Biosciences).

Immunoprecipitation

ITGA2B or ITGA2B-ceC96–specific antibodies (0.1 mg/ml) and protein samples (300 μg) from PBMCs or 293T cell lysates were used for immunoprecipitation and coimmunoprecipitation assays according to the manufacturer's protocol (Immunoprecipitation Kit, Thermo Scientific, 26147; Co-Immunoprecipitation Kit, Thermo Scientific, 26149).

Apoptosis and cell cycle analysis

Apoptotic cells and the cell cycle were detected with the fluorescein isothiocyanate (FITC) Annexin V Apoptosis Detection Kit with propidium iodide (BioLegend, 640914) and a cell cycle detection kit (Keygen Biotech, KGA512) according to the manufacturer's protocols.

Immunogenicity prediction

The SYFPEITHI database was used to identify and score candidate carboxyethylated peptides, and the scores were based on the original cysteine sequence, not the modified one. HLA-B*2705, HLA-B*2709, HLA-DRB1*01, HLA-DRB1*03, and HLA-DRB1*04 were scored, and all candidate peptides were sorted by the average score for the five HLA subtypes. A CLIP peptide (87-101) was used as a negative control peptide.

*HLA-DRA*01/HLA-DRB1*04 molecule complex construction and expression*

To produce the HLA-DRA*01:01 (IMGT/HLA accession no. HLA00662)/HLA-DRB1*04:02:01 (IMGT/HLA accession no. HLA00687) heterodimer, cDNA sequences of the extracellular domains were codon optimized and introduced into the pFastBac Dual plasmid. A nucleotide segment was introduced to link the CLIP peptide (PVSKMRMATPLLMQA) to the N terminus of the DRB1 chain through a 15–amino acid long linker that includes a thrombin cleavage site. A strep II tag was added before the CLIP peptide. The C-terminal part of the beta chain contained a Cys followed by a human rhinovirus (HRV) 3C protease cleavage site, a basic leucine zipper, a FLAG tag, and a BirA biotinylation site. The C-terminal region of the alpha chain also contained a cysteine residue, an HRV 3C protease cleavage site, an acidic leucine zipper, and a His tag. An interchain disulfide bond was formed by the C-terminally introduced Cys residues, and the leucine zipper pair promoted MHC dimer formation, whereas the His and FLAG tags facilitated protein purification. The sequences encoding the extracellular domains of HLA-DM were also inserted into the two separated MCSs in the pFastBac Dual plasmid. A strep tag and 6His-tag preceded by a FLAG tag were added to the C terminus of DMA and DMB, respectively. HLA was produced with a baculovirus expression system according to the manufacturer's instructions. In brief, recombinant Bacmid DNA carrying the HLA-DR sequence was transfected into Sf9 insect cells, and high titers of baculovirus were obtained after three rounds of viral amplification. The baculovirus was then used to infect the High5 cell line grown in suspension in serum-free medium for secretory expression. Supernatants harvested on day 3 or 4 were buffer exchanged into equilibrium buffer (TBS, pH 8.0), concentrated using the TFF system, and subjected to affinity purification using anti-FLAG resin (Genscript). Buffer exchange (PBS) and concentration of eluates were conducted by centrifugation (6000*g* for 4 hours at 4°C), and purity was confirmed by SDS–polyacrylamide gel electrophoresis.

*In vitro HLA-DRA*01/HLA-DRB1*04 peptide exchange assays*

We first constructed the alpha-beta complex of HLA-DR and HLA-DM with different labels. After testing the thrombin efficiency, HLA-DR (1 μM) was incubated with thrombin at 37°C for 1 hour for thrombin cleavage. Then, the ceC96-27 and wtC96-27 peptides (27 amino acids) were added to the exchange system (37°C for 2 hours with or without HLA-DM). After the reaction, His pull-down experiments were performed to remove unbound ceC96-27 (1 μM) and wtC96-27 (1 μM) and the Strep-CLIP peptide was exchanged. Finally, anti-ITGA2B-ceC96 (0.2 μg/ml), anti-Strep (1:1000), and anti-FLAG (1:1000) antibodies were used to verify ceC96-27 exchange by a dot blot assay.

FP assays

A FAM-labeled Epstein-Barr virus EBNA1 peptide (482-496, FAM-KGGGAEGLRALLARSHVER) was used as a prebinding peptide (FAM-prepeptide) for the HLA-DRA*01/HLA-DRB1*04 complex. For peptide competition assays, a fluorescent HLA-DRA*01/HLA-DRB1*04-FAM-prepeptide complex was generated by incubation of HLA-DRA*01/HLA-DRB1*04/CLIP (1 μM) and 500 nM HLA-DM with 500 nM FAM-prepeptide for 3 hours at 37°C in sodium citrate buffer (150 mM sodium chloride, 50 mM sodium citrate buffer, pH 5.2). Unbound peptide was later removed with a gel filtration column (PD10; GE Healthcare).

Competing peptides were added to the HLA-DRA*01/HLA-DRB1*04-FAM-prepetide complex (typically in a 100-μl volume in 96-well plates) for the indicated times and concentrations. The dissociation of FAM prepeptide from the HLA-DRA*01/HLA-DRB1*04-FAM-prepeptide complex (100 nM, with 30 nM HLA-DM) was measured using an Infinite F200 plate reader (Tecan) with a 485/20 nm band-pass and 535/25 nm band-pass filter set. FP values were read at 25°C.

HLA-peptide structure analysis

The structure of HLA-DRA*01/HLA-DRB1*04:01/α-Enolase 26 (PDB: 5LAX) was used as the starting template. Residues in the beta chain of HLA-DRB1*04:02:01, which are different from those of HLA-DRB1*04:01, were mutated to be identical to those in HLA-DRB1*04:02:01. The peptide SKGLFRAAVPSGAS bound to HLA-DRA*01/HLA-DRB1*04:01 in the crystal structure was mutated to FLCPWRAEGGQCPS. The new model obtained in this way was further optimized by structure minimization in UCSF Chimera (*20*). Most parameters were set to the default values, excluding the steepest descent steps, which were set to 1000, and the conjugate gradient steps were set to 100. The interactions between HLA and peptides were analyzed by PISA. All the structure figures were generated with PyMol (*47*). The structure of HLA-DRA*01/HLA-DRB1*04:02:01 in complex with the carboxyethylated peptide was also edited in UCSF Chimera and optimized with structure minimization using the same protocol.

HLA haplotype assays

The HLA haplotype identification report for patients with AS was provided by Beijing Bofurui Gene Diagnostics Ltd. Briefly, the HLA haplotype included the identification of HLA-A, HLA-D, and HLA-DRB1 alleles.

DC culture and identification

DC culture and identification were performed as previously described (*48*). Briefly, PBMCs were isolated and then resuspended in AIM-V medium. Immature DC (iDC) medium (AIM-V medium containing 800 U/ml of granulocyte-macrophage colony-stimulating factor (GM-CSF) and 500 U/ml of IL-4) was added to the adherent cells, which were incubated in a 37°C and 5% CO_2 incubator. On day 6 of iDC culture, an equal volume of mDC medium was added (AIM-V medium containing of 1600 U/ml of GM-CSF, 1000 U/ml of IL-4, and 10 ng/ml of TNF-α, 10 ng/ml of IL-1β, 320 ng/ml of IL-6, and 2 μg/ml of PGE2) with the peptides (wtC96 an ceC96) to achieve a final concentration of 10 μg/ml. The ratio of the original conditioned medium to the mDC medium was 1:1 for 16 to 18 hours. iDCs and mDCs were primarily identified through flow cytometry. The dilution ratios and clone names of the commercial antibodies used for flow cytometry are listed in table S6.

T cell stimulation assays

mDCs and T cells from the same individual were cocultured at a 1:10 ratio in RPMI-1640 medium supplemented with 10 ng/ml of IL-7, 10% FBS, and 1% penicillin-streptomycin on day 0. After 3 days, 50 U/ml of IL-2 was added and the cells were cultured for 7 days. On day 10, mDCs were added to the T cells at a 1:10 ratio for restimulation for 8 hours in the presence of the blocking agents brefeldin A (5 μg/ml) and monensin (2 μM) (*48*). Cells expressing CD137, IFN-γ, IL-2, and TNF-α were detected by flow cytometry. After another 10 days of coculture (day 20), $CD4^+$ and $CD8^+$ T cells were sorted and IFN-γ levels were measured by ELISpot assays (DAKEWE, 2110005) according to the manufacturer's protocol. After an additional 10 days of coculture (day 30), $CD4^+$ and $CD8^+$ T cells were sorted and single-cell TCR sequencing was performed. For positive controls, T cells were stimulated with phorbol 12-myristate 13-acetate (PMA), PMA + ionomycin, or pooled peptide-treated mDCs.

Single-cell TCR repertoire analysis

$CD4^+$ and $CD8^+$ T cells obtained from PBMCs stimulated with DCs (DC-NC and DC-ceC96) for 30 days were isolated by FACS (BD FACSAria III). $CD3^+CD4^+$ T cells and $CD3^+CD8^+$ T cells were collected, and 1×10^6 T cells/ml were used for library preparation. 10X Genomics Chromium Single Cell V(D)J immune repertoire profiling was performed and analyzed by Novogene (*49*). TCR clone subtype analysis was then performed using Cell Ranger (10X Genomics) and VDJtools.

TCR validation

TCR validation experiments were performed as previously described (*17*, *18*). Briefly, candidate TCRα-TCRβ pairs were cloned into pCDH-MSCV-MCS-EF1-copGFP-T2A-Puro vectors and retrovirally expressed in $CD4^+$ TCR-free Jurkat cells. The HLA-DRB1*04/ceC96 tetramer was used to validate the candidate TCRs. The levels of IL-2 secreted from DC-pulsed TCR Jurkat cells were measured with an IL-2 ELISA kit (DAKEWE, 1110202).

HLA-DR mice

Abb knockout/transgenic HLA-DR4 mice (B6.129S2-H2-Ab1^{tm1Gru} Tg (HLA-DRA/H2-Fa, HLA-DRB1*0401/H2-Eb) 1Kito) were purchased from Taconic and ethics approval was granted by the Ethics Committee of Fourth Military Medical University.

In vivo disease models

An in vivo mouse disease model was generated by subcutaneously injecting 400 μg of peptide (ITGA2B-wtC96 or ITGA2B-ceC96) emulsified in CFA containing 250 μg of H37Ra *Mycobacterium tuberculosis*. Fourteen days later, the mice were boosted with a second subcutaneous injection of the same amount of antigen emulsified in IFA. On day 21, the mice were boosted with a third subcutaneous injection of the same amount of antigens with IFA. On days 0, 7, 14, and 21, peripheral blood was collected before immunization. The mice were sacrificed on day 31 and samples were analyzed. C57BL/6N mice were immunized similarly as a control. In the KLH-coupled immunization groups, 1 mM 3-HPA (pH = 7, neutralized with sodium hydroxide) was added to the water 28 days after first KLH-coupled ITGA2B-wtC96 or ITGA2B-ceC96 immunization in HLA-DRB1*04 mice.

To evaluate the general colitis model, the disease activity index was determined from three scores (*50*). Colonic histopathology scoring was conducted in a double-blinded fashion by at least four independent scorers following a previously reported protocol (*24*, *51*). Changes in fecal contents were measured by lipocalin-2 ELISAs and fecal occult blood tests. H&E, periodic acid–Schiff, and Safranin O–Fast Green staining were performed as previously described (*52*, *53*).

To evaluate bone mass, a micro-computed tomography (micro-CT) system (GE Healthcare, USA) was used as reported previously (*54*). Mice were sacrificed, and specimens were fixed with 10% (vol/vol) paraformaldehyde for 24 hours and then imaged using a micro-CT scanner at a resolution of 6.5 μm. Three-dimensional reconstruction and structural parameter quantification were performed with Micview V2.1.2 software. Segmentation thresholds of 200 to 500 mg of hemagglutinin (HA)/cm^3 were used to identify the newly formed bone, which was anticipated to exhibit a lower bone mineral density (BMD), and thresholds of 550 to 3000 mg of HA/cm^3 were applied to measure normal mature bone with a higher BMD (*6*).

Detection of anti–ceC96-antibody in plasma by ELISA

An ELISA protocol was established to measure anti-ceC96 antibody levels in plasma. Three peptides, wtC96-linear (VFLCPWRAEGGQCPSL-LFDLRDETRNV), ceC96-linear (VFLCPWRAEG-GQC$_{ce}$PSLLFDLRDETRNV), and ceC96-bridge (VFLCPWRAEGGQC$_{ce}$PSLLFDLRDCTRNV, Cys4 & Cys23 formed a disulfide bridge) were used to determine anti-ceC96 and anti-wtC96 antibody levels. First, ELISA plates were separately precoated with the three peptides (10 μg/ml) at 4°C overnight. Then, rabbit anti-ceC96 and anti-wtC96 antibodies were added at different concentrations (100, 10, or 1 ng/ml and 100, 10, or 1 pg/ml). Rabbit IgG (1 μg/ml) was used as a negative control. ELISAs were then performed following a standard procedure (*55*) to clarify the specificity and effectiveness

of the ceC96-bridge peptide. To detect anti-ceC96 antibody levels in plasma from humans and mice, ceC96-bridge peptide was coated on the plates to directly interact with sample sera and followed by incubating with HRP-conjugated secondary anti-human or anti-mouse antibody to obtain the optical density values. HC mixtures (HC_{mix}) of healthy donors in group 2 (table S1) were used as a relatively quantitative standard in different batches of anti-ceC96 antibody assays, whereas human anti-ITGA2B-ceC96–specific autoantibodies were calibrated by HC_{mix}.

Tetramer construction

The HLA-DRA*01/HLA-DRB1*04 molecular complex was designed with a BirA biotinylation site in the C-terminal region of the beta chain. After construction and purification of the HLA-DRA*01/HLA-DRB1*04-CLIP (negative control), HLA-DRA*01/HLA-DRB1*04-influenza A (positive control), HLA-DRA*01/HLA-DRB1*04-wtC96, and HLA-DRA*01/HLA-DRB1*04-ceC96 complexes, biotinylation was performed to generate peptide-MHC monomers. These peptide-MHC monomers were biotinylated to enable multimerization of tetrameric complexes on streptavidin. Furthermore, allophycocyanin-streptavidin and phycoerythrin-streptavidin were used for flow cytometry.

Metabolic profiling

Metabolic profiling was performed (*56*) by Metabo-Profile Biotechnology (Shanghai) Co., Ltd. Briefly, standards for all targeted metabolites obtained from Sigma-Aldrich (St. Louis, MO, USA), Steraloids Inc. (Newport, RI, USA) and TRC Chemicals (Toronto, Ontario, Canada) were accurately weighed and prepared at a concentration of 5.0 mg/ml. After derivatization, the samples were transferred to a new 96-well plate with 10 μl of internal standards in each well. An ultraperformance liquid chromatography coupled to tandem mass spectrometry (UPLC–MS/MS) system (ACQUITY UPLC-Xevo TQ-S, Waters Corp., Milford, MA, USA) was used to quantitate all targeted metabolites. Three types of quality control samples (i.e., test mixtures, internal standards, and pooled biological samples) were used. The raw data files generated by UPLC–MS/MS were processed using MassLynx software (v4.1, Waters, Milford, MA, USA) to perform peak integration, calibration, and quantitation.

Statistical analysis

Error bars represent the SEM or SD, as indicated in the figure legends. Statistical significance was determined using Prism version 8.0 software (GraphPad Software, CA, USA). Differences were deemed significant at $P < 0.05$. Two-way or one-way ANOVA followed by Dunnett's post-test, Tukey's multiple-comparisons test, or the Kruskall–Wallis test (for subgroup analyses) was performed for multiple comparisons, and Student's t test was performed for other experiments to compare mean values. $^{****}P < 0.0001$; $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$; and ns, not significant ($P > 0.05$).

REFERENCES AND NOTES

1. E. S. Golub, Immune tolerance and autoreactivity: Do they point to a regulatory role of the immune system for normal cell function? *Cell* **27**, 417–418 (1981). doi: 10.1016/0092-8674(81)90382-2; pmid: 6101196
2. L. Van Parijs, A. K. Abbas, Homeostasis and self-tolerance in the immune system: Turning lymphocytes off. *Science* **280**, 243–248 (1998). doi: 10.1126/science.280.5361.243; pmid: 9535647
3. D. Ummarino, ACPA status influences RA development. *Nat. Rev. Rheumatol.* **13**, 450 (2017). doi: 10.1038/nrrheum.2017.111; pmid: 28660910
4. X. Dong *et al*., ACPA mediates the interplay between innate and adaptive immunity in rheumatoid arthritis. *Autoimmun. Rev.* **17**, 845–853 (2018). doi: 10.1016/j.autrev.2018.02.014; pmid: 30007856
5. E. Gracey *et al*., Tendon and ligament mechanical loading in the pathogenesis of inflammatory arthritis. *Nat. Rev. Rheumatol.* **16**, 193–207 (2020). doi: 10.1038/s41584-019-0364-x; pmid: 32080619
6. M. N. van Tok *et al*., Interleukin-17A inhibition diminishes inflammation and new bone formation in experimental spondyloarthritis. *Arthritis Rheumatol.* **71**, 612–625 (2019). doi: 10.1002/art.40770; pmid: 30390386
7. N. Kanda, K. Takeda, S. Hatakeyama, M. Matsumura, Ankylosing spondylitis presenting with enthesitis at an uncommon site and fever of unknown origin. *BMJ Case Rep.* **12**, e230113 (2019). doi: 10.1136/bcr-2019-230113; pmid: 31401572
8. H. Mielants, F. De Keyser, D. Baeten, F. Van den Bosch, Gut inflammation in the spondyloarthropathies. *Curr. Rheumatol. Rep.* **7**, 188–194 (2005). doi: 10.1007/s11926-996-0038-y; pmid: 15918994
9. D. Mauro *et al*., Ankylosing spondylitis: An autoimmune or autoinflammatory disease? *Nat. Rev. Rheumatol.* **17**, 387–404 (2021). doi: 10.1038/s41584-021-00625-y; pmid: 34113018
10. J. J. Calvete, A. Henschen, J. González-Rodríguez, Complete localization of the intrachain disulphide bonds and the N-glycosylation points in the alpha-subunit of human platelet glycoprotein IIb. *Biochem. J.* **261**, 561–568 (1989). doi: 10.1042/bj2610561; pmid: 2775232
11. E. W. Miles, J. P. Kraus, Cystathionine beta-synthase: Structure, function, regulation, and location of homocystinuria-causing mutations. *J. Biol. Chem.* **279**, 29871–29874 (2004). doi: 10.1074/jbc.R400005200; pmid: 15087459
12. M. Meier, M. Janosik, V. Kery, J. P. Kraus, P. Burkhard, Structure of human cystathionine beta-synthase: A unique pyridoxal 5′-phosphate-dependent heme protein. *EMBO J.* **20**, 3910–3916 (2001). doi: 10.1093/emboj/20.15.3910; pmid: 11483494
13. J. Ereño-Orbea, T. Majtan, I. Oyenarte, J. P. Kraus, L. A. Martínez-Cruz, Structural insight into the molecular mechanism of allosteric activation of human cystathionine β-synthase by S-adenosylmethionine. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E3845–E3852 (2014). doi: 10.1073/pnas.1414545111; pmid: 25197074
14. J. Zhu *et al*., Structure of a complete integrin ectodomain in a physiologic resting state and activation and deactivation by applied forces. *Mol. Cell* **32**, 849–861 (2008). doi: 10.1016/j.molcel.2008.11.018; pmid: 19111664
15. H. Rammensee, J. Bachmann, N. P. Emmerich, O. A. Bachor, S. Stevanović, SYFPEITHI: Database for MHC ligands and peptide motifs. *Immunogenetics* **50**, 213–219 (1999). doi: 10.1007/s002510050595; pmid: 10602881
16. J. D. Reveille *et al*., HLA class I and II alleles in susceptibility to ankylosing spondylitis. *Ann. Rheum. Dis.* **78**, 66–73 (2019). doi: 10.1136/annrheumdis-2018-213779; pmid: 30341055
17. D. B. Keskin *et al*., Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature* **565**, 234–239 (2019). doi: 10.1038/s41586-018-0792-9; pmid: 30568305
18. Z. Hu *et al*., A cloning and expression system to probe T-cell receptor specificity and assess functional avidity to neoantigens. *Blood* **132**, 1911–1921 (2018). doi: 10.1182/blood-2018-04-843763; pmid: 30150207
19. S. Pan, T. Trejo, J. Hansen, M. Smart, C. S. David, HLA-DR4 (DRB1*0401) transgenic mice expressing an altered CD4-binding site: Specificity and magnitude of DR4-restricted T cell response. *J. Immunol.* **161**, 2925–2929 (1998). doi: 10.4049/jimmunol.161.6.2925; pmid: 9743354
20. E. F. Pettersen *et al*., UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254
21. C. Gerstner *et al*., Functional and structural characterization of a novel HLA-DRB1*04:01-restricted α-enolase T cell epitope in rheumatoid arthritis. *Front. Immunol.* **7**, 494 (2016). doi: 10.3389/fimmu.2016.00494; pmid: 27895642
22. K. Ito *et al*., HLA-DR4-IE chimeric class II transgenic, murine class II-deficient mice are susceptible to experimental allergic encephalomyelitis. *J. Exp. Med.* **183**, 2635–2644 (1996). doi: 10.1084/jem.183.6.2635; pmid: 8676084
23. G. E. Hammer *et al*., Expression of A20 by dendritic cells preserves immune homeostasis and prevents colitis and spondyloarthritis. *Nat. Immunol.* **12**, 1184–1193 (2011). doi: 10.1038/ni.2135; pmid: 22019834
24. T. Sujino *et al*., Tissue adaptation of regulatory and intraepithelial $CD4^+$ T cells controls gut inflammation. *Science* **352**, 1581–1586 (2016). doi: 10.1126/science.aaf3892; pmid: 27256884
25. J. Shimizu *et al*., Propionate-producing bacteria in the intestine may associate with skewed responses of IL10-producing regulatory T cells in patients with relapsing polychondritis. *PLOS ONE* **13**, e0203657 (2018). doi: 10.1371/journal.pone.0203657; pmid: 30235279
26. J. M. Zhao, X. Chen, K. Cheng, Q. Shi, K. Peng, Anserine and glucosamine supplementation attenuates the levels of inflammatory markers in rats with rheumatoid arthritis. *AMB Express* **10**, 57 (2020). doi: 10.1186/s13568-020-00987-8; pmid: 32198574
27. A. Duscha *et al*., Propionic acid shapes the multiple sclerosis disease course by an immunomodulatory mechanism. *Cell* **180**, 1067–1080.e16 (2020). doi: 10.1016/j.cell.2020.02.035; pmid: 32160527
28. T. P. Nguyen-Vo *et al*., Development of 3-hydroxypropionic-acid-tolerant strain of Escherichia coli W and role of minor global regulator yieP. *Metab. Eng.* **53**, 48–58 (2019). doi: 10.1016/j.ymben.2019.02.001; pmid: 30738917
29. X. Li, L. Chen, X. Wang, P. Tian, Physiological investigations of the influences of byproduct pathways on 3-hydroxypropionic acid production in *Klebsiella pneumoniae*. *J. Basic Microbiol.* **59**, 1195–1207 (2019). doi: 10.1002/jobm.201800640; pmid: 31617952
30. R. Y. Ji *et al*., Metabolic engineering of yeast for the production of 3-hydroxypropionic acid. *Front. Microbiol.* **9**, 2185 (2018). doi: 10.3389/fmicb.2018.02185; pmid: 30298059
31. F. Humphries *et al*., Succination inactivates gasdermin D and blocks pyroptosis. *Science* **369**, 1633–1637 (2020). doi: 10.1126/science.abb9818; pmid: 32820063
32. A. E. Lepack *et al*., Dopaminylation of histone H3 in ventral tegmental area regulates cocaine seeking. *Science* **368**, 197–201 (2020). doi: 10.1126/science.aaw8806; pmid: 32273471
33. S. Bournazos *et al*., Antibody fucosylation predicts disease severity in secondary dengue infection. *Science* **372**, 1102–1105 (2021). doi: 10.1126/science.abc7303; pmid: 34083490
34. D. Zhang *et al*., Metabolic regulation of gene expression by histone lactylation. *Nature* **574**, 575–580 (2019). doi: 10.1038/s41586-019-1678-1; pmid: 31645732
35. S. van der Linden, H. A. Valkenburg, A. Cats, Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis Rheum.* **27**, 361–368 (1984). doi: 10.1002/art.1780270401; pmid: 6231933
36. F. C. Arnett *et al*., The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* **31**, 315–324 (1988). doi: 10.1002/art.1780310302; pmid: 3358796
37. M. C. Hochberg, Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* **40**, 1725 (1997). doi: 10.1002/art.1780400928; pmid: 9324032
38. Z. Zhao *et al*., LC-MS/MS analysis unravels deep oxidation of manganese superoxide dismutase in kidney cancer. *Int. J. Mol. Sci.* **18**, 319 (2017). doi: 10.3390/ijms18020319; pmid: 28165386
39. Y. Jiang *et al*., Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature* **567**, 257–261 (2019). doi: 10.1038/s41586-019-0987-8; pmid: 30814741
40. J. Chen *et al*., Open search-based proteomics reveals widespread tryptophan modifications associated with hypoxia in lung cancer. *Oxid. Med. Cell. Longev.* **2022**, 2590198 (2022). doi: 10.1155/2022/2590198; pmid: 35535361

41. J. R. Yates 3rd, Pivotal role of computers and software in mass spectrometry - SEQUEST and 20 years of tandem MS database searching. *J. Am. Soc. Mass Spectrom.* **26**, 1804–1813 (2015). doi: 10.1007/s13361-015-1220-0; pmid: 26286455
42. S. Na, E. Paek, Software eyes for protein post-translational modifications. *Mass Spectrom. Rev.* **34**, 133–147 (2015). doi: 10.1002/mas.21425; pmid: 24889695
43. M. Bern, Y. J. Kil, C. Becker, Byonic: advanced peptide and protein identification software. *Curr. Protoc. Bioinformatics.* **Chapter 13**, 13.20.1–13.20.14 (2012). doi: 10.1002/0471250953.bi1320s40; pmid: 23255153
44. C. Fraley, A. E. Raftery, How many clusters? Which clustering method? Answers via modelbased cluster analysis. *Comput. J.* **41**, 578–588 (1998). doi: 10.1093/comjnl/41.8.578
45. M. Y. Zhang *et al*., Disrupting CD147-RAP2 interaction abrogates erythrocyte invasion by *Plasmodium falciparum*. *Blood* **131**, 1111–1121 (2018). doi: 10.1182/blood-2017-08-802918; pmid: 29352039
46. Y. Zhai *et al*., TNFAIP3-DEPTOR complex regulates inflammasome secretion through autophagy in ankylosing spondylitis monocytes. *Autophagy* **14**, 1629–1643 (2018). doi: 10.1080/15548627.2018.1458804; pmid: 29940800
47. W. L. DeLano, "Pymol: An open-source molecular graphics tool" (CCP4 Newsletter on Protein Crystallography, 2002); https://legacy.ccp4.ac.uk/newsletters/newsletter40/11_pymol.html.
48. Q. T. Wang *et al*., Tumor-associated antigen-based personalized dendritic cell vaccine in solid tumor patients. *Cancer Immunol. Immunother.* **69**, 1375–1387 (2020). doi: 10.1007/s00262-020-02496-w; pmid: 32078016
49. Y. Luo *et al*., Single-cell transcriptomic analysis reveals disparate effector differentiation pathways in human T_{reg} compartment. *Nat. Commun.* **12**, 3913 (2021). doi: 10.1038/s41467-021-24213-6; pmid: 34162888
50. L. Gao *et al*., A resident stromal cell population actively restrains innate immune response in the propagation phase of colitis pathogenesis in mice. *Sci. Transl. Med.* **13**, eabb5071 (2021). doi: 10.1126/scitranslmed.abb5071; pmid: 34290057
51. T. E. Adolph *et al*., Paneth cells as a site of origin for intestinal inflammation. *Nature* **503**, 272–276 (2013). doi: 10.1038/nature12599; pmid: 24089213
52. S. Shim *et al*., Rebamipide ameliorates radiation-induced intestinal injury in a mouse model. *Toxicol. Appl. Pharmacol.* **329**, 40–47 (2017). doi: 10.1016/j.taap.2017.05.012; pmid: 28526636
53. M. D. Newton *et al*., Longitudinal characterization of intervertebral disc remodeling following acute annular injury in a rat model of degenerative disc disease. *Connect. Tissue Res.* **61**, 568–576 (2020). doi: 10.1080/03008207.2019.1635589; pmid: 31232119
54. M. L. Bouxsein *et al*., Guidelines for assessment of bone microstructure in rodents using micro-computed tomography. *J. Bone Miner. Res.* **25**, 1468–1486 (2010). doi: 10.1002/jbmr.141; pmid: 20533309
55. K. Wang *et al*., CD147-spike protein is a novel route for SARS-CoV-2 infection to host cells. *Signal Transduct. Target. Ther.* **5**, 283 (2020). doi: 10.1038/s41392-020-00426-x; pmid: 33277466
56. G. Xie *et al*., A metabolite array technology for precision medicine. *Anal. Chem.* **93**, 5709–5717 (2021). doi: 10.1021/acs.analchem.0c04686; pmid: 33797874
57. M. Lü, B. Xia, Polymorphism of HLA-DRB1 gene shows no strong association with ulcerative colitis in Chinese patients. *Int. J. Immunogenet.* **33**, 37–40 (2006). doi: 10.1111/j.1744-313X.2005.00559.x; pmid: 16426241
58. P. Xiong *et al*., Lack of association between HLA-A, -B and -DRB1 alleles and the development of SARS: A cohort of 95 SARS-recovered individuals in a population of Guangdong, southern China. *Int. J. Immunogenet.* **35**, 69–74 (2008). doi: 10.1111/j.1744-313X.2007.00741.x; pmid: 18186801
59. Z. Li *et al*., HLA polymorphisms are associated with *Helicobacter pylori* infected gastric cancer in a high risk population, China. *Immunogenetics* **56**, 781–787 (2005). doi: 10.1007/s00251-004-0723-9; pmid: 15650879

ACKNOWLEDGMENTS

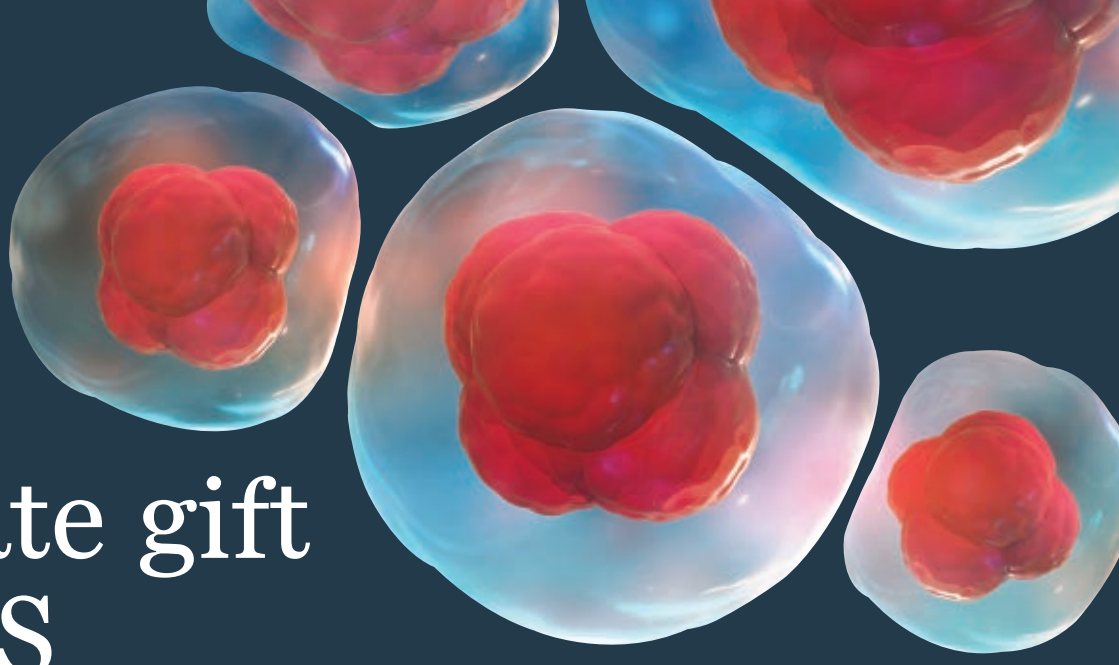
We thank B.-Q. Jin and J. Liang for providing insightful suggestions, J.-L. Xia for guidance on selecting an appropriate statistical test and for sufficient powering of experiments, S. Zhang for timely efforts in the handling of related documents, and J. Liang for critical feedback on the manuscript. **Funding:** This work was supported by National Basic Research Program of China (grant 2015CB553700 to P.Z. and Z.-N.C.), the National Natural Science Foundation of China (grant 32070157 to J.-H.Y., grant 82101885 to Z.Y., and 2021 National Natural Science Foundation Outstanding Youth Science Fund Project to L.C.), the Department of Science and Technology of Henan Province (grant 201400210500), Xinfa Pharmaceuticals of Shandong Province (Tai-Shan Scholarship to J.-H.Y.), Young Elite Scientists Sponsorship Program by CAST (YESS grant 2019QNRC001 to Z.Y.). **Author contributions:** Conceptualization: P.Z., Z.-N.C., J.-H.Y., Y.Z., and L.C.; Formal analysis: Y.Z., J.-H.Y., X.C., P.L., H.-Y.L., X.-M.Y., L.C., P.Z., and Z.-N.C.; Funding acquisition: P.Z., Z.-N.C., J.-H.Y., Y.Z., and L.C.; Investigation: Y.Z., Q.Z., Z.-H.Z., H.-J.B., X.Y., X.C., P.L., H.-Y.L., P.Z., L.C., R.C., L.-N.F., H.-Y.S., Kun Z., J.-S.Z., S.-R.C., S.-Y.C., T.Z., J.-J.C., B.W., X.-X.S., Z.F., Y.-M.Z., L.-J.W., X.-M.L, Q.H., Y.W., Y.C., G.N., L.Y., R.-H.X., Kui Z., R.Z., and B.Z.; Methodology: Y.Z., J.-H.Y., X.C., P.L., and L.C.; Supervision: P.Z. and Z.-N.C.; Writing – original draft: Y.Z., J.-H.Y., L.C., P.Z., and Z.-N.C.; Writing – review & editing: Y.Z., J.-H.Y., L.C., P.Z., Z.-N.C., and H.-J.B. **Competing interests:** P.Z., Z.-N.C, Y.Z., L.C., J.-H.Y., H.-J.B., X.-M.Y., and Z.Z. are inventors on patents applied for by Fourth Military Medical University related to cysteine carboxyethylation. The remaining authors declare no competing interests. **Data and materials availability:** The raw RNA-seq data used and generated in this study have been deposited in the GEO database under accession number GSE207037 and in the human GSA database under accession number HRA002493. All other data in this paper are presented in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg2482
Figs. S1 to S22
Tables S1 to S6
External Databases S1 and S2
MDAR Reproducibility Checklist

Submitted 20 December 2020; resubmitted 23 December 2021
Accepted 31 January 2023
10.1126/science.abg2482

1848 SOCIETY
AAAS



An estate gift to AAAS

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

Today, we are dedicated to advocating for science and scientific evidence to be fully and positively integrated into public policy and for the community to speak with one voice to advance science and engineering in the United States and around the world.

By making AAAS a beneficiary of your will, trust, retirement plan or life insurance policy, you will become a member of our 1848 Society and will help fuel our work on behalf of science and society – including publishing the world's most promising, innovative research in the *Science* family of journals and engaging in the issues that matter locally, nationally and around the world.

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

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

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

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

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

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

cut here

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

1848 SOCIETY
AAAS

PLEASE CONTACT ME AT:

Name: _____

Address: _____

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

Email: _____ Phone: _____

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

RESEARCH ARTICLES

STRUCTURAL BIOLOGY

Structures of BIRC6-client complexes provide a mechanism of SMAC-mediated release of caspases

Moritz Hunkeler[1,2], Cyrus Y. Jin[1,2], Eric S. Fischer[1,2]*

Tight regulation of apoptosis is essential for metazoan development and prevents diseases such as cancer and neurodegeneration. Caspase activation is central to apoptosis, and inhibitor of apoptosis proteins (IAPs) are the principal actors that restrain caspase activity and are therefore attractive therapeutic targets. IAPs, in turn, are regulated by mitochondria-derived proapoptotic factors such as SMAC and HTRA2. Through a series of cryo–electron microscopy structures of full-length human baculoviral IAP repeat–containing protein 6 (BIRC6) bound to SMAC, caspases, and HTRA2, we provide a molecular understanding for BIRC6-mediated caspase inhibition and its release by SMAC. The architecture of BIRC6, together with near-irreversible binding of SMAC, elucidates how the IAP inhibitor SMAC can effectively control a processive ubiquitin ligase to respond to apoptotic stimuli.

Genetically encoded cell death programs such as apoptosis, necroptosis, and pyroptosis remove infected, damaged, or obsolete cells during development and are essential in all metazoans (*1*). Aberrant activity or lack of regulation of cell death pathways leads to a wide array of human pathologies such as neurodegeneration, autoinflammatory disease, and cancer (*2*, *3*). Apoptotic signaling pathways converge on activation of cysteine-dependent aspartate-directed proteases (caspases) to trigger execution of the apoptotic program (*4*). Inhibitor of apoptosis proteins (IAPs) keep cell death commitment in check by direct binding and inhibition of initiator caspase-9 (casp-9) and executioner caspases-3 and -7 (casp-3 and casp-7) (*5*). Many cancer cells express elevated levels of IAPs and have heightened apoptotic thresholds to resist apoptotic signals and cytotoxic therapies (*6*). In mammals, seven IAPs, also referred to as BIRCs [baculoviral IAP repeat (BIR) domain–containing proteins] exist (*5*, *7*, *8*). The BIR domain found in all IAPs interacts with the conserved IAP binding motif (IBM) of caspases (*9*) (fig. S1A). Most IAPs also act as ubiquitin ligases mediating polyubiquitylation and thereby degradation of caspases (*5*).

Apoptotic stimuli lead to BCL-2 family–mediated release of proapoptotic molecules from the mitochondria, including cytochrome c (*10*), apoptosis-inducing factor (*11*), second mitochondria-derived activator of caspases [SMAC, also known as direct IAP-binding protein with low isoelectric point (DIABLO)] (*12*, *13*), and the serine protease high temperature requirement protein A2 (HTRA2) (*14*). Similar to caspases, SMAC and HTRA2 both directly interact with BIR domains of IAPs through N-terminal IBMs (*9*) (fig. S1A), thereby blocking IAP-mediated caspase inhibition and releasing the brakes on cell death (*5*, *14*). These observations led to the development of small-molecule SMAC mimetics that target the IAP–caspase interaction (*15*), several of which are currently undergoing clinical exploration (*16*).

BIRC6 [also called APOLLON (*17*) or BRUCE (*18*)] plays critical roles in cell division (*19*, *20*) and regulation of autophagy (*21*, *22*) and was shown to exhibit prototypical anti-apoptotic activity (*17*, *18*, *23*–*26*). The *Drosophila* BIRC6 homolog dBRUCE has been implicated in inhibiting apoptosis by targeting Reaper (*27*, *28*), a functional homolog of SMAC with high sequence similarity in its IBM (*29*). BIRC6 is an essential IAP (*24*) and is evolutionarily the oldest IAP to acquire E3 ubiquitin ligase activity (*30*). The large multidomain 530-kDa protein has a single BIR domain close to the N terminus and an E2-E3 hybrid ubiquitin-conjugating (UBC) domain close to the C terminus, whereas the rest of the protein remains largely uncharacterized (Fig. 1A). BIRC6 interacts with and inhibits casp-3, casp-7, and casp-9, and these interactions are inhibited by SMAC (*23*, *31*). BIRC6 has also been established as a substrate for caspases as well as HTRA2 (Fig. 1B) (*23*).

Owing to a lack of structural information, how BIRC6 binds to and inhibits proapoptotic caspases and how this inhibition is effectively counteracted by SMAC and HTRA2 remains elusive. We present cryo–electron microscopy (cryo-EM) structures of full-length BIRC6 alone or bound to SMAC, casp-3, casp-7, or HTRA2 and discuss how BIRC6 engages diverse substrates and inhibitors to control apoptosis. We find that a dimeric architecture of BIRC6 establishes an accommodating central cavity that allows for competitive binding of diverse factors recruited through perfectly spaced BIR domains and stabilized by electrostatic interactions. This architecture, together with the near-irreversible binding of SMAC, reconciles how caspase inhibition and ubiquitylation by BIRC6 is released upon apoptosis.

BIRC6 is a UBA6-dependent E2/E3 chimera

BIRC6 exhibits UBC domain-dependent anti-apoptotic activity and binds to caspases, HTRA2, and SMAC through its BIR domain in cells (*23*, *25*, *26*, *32*). To reconstitute these activities in a fully recombinant system and to enable structural studies, we expressed and purified full-length BIRC6 (fig. S1B). By examining the Cancer Dependency Map [DepMap (*33*)], we noticed a strong correlation between BIRC6 and the noncanonical ubiquitin-activating enzyme UBA6. On the basis of this observation, we set out to test whether BIRC6 prefers UBA6 over the canonical ubiquitin E1 enzyme UBA1. Auto-ubiquitylation assays using UBA1 or UBA6 established a clear preference for UBA6 with negligible activity observed with UBA1 (Fig. 1C), consistent with previous observations that UBA6 and BIRC6 cooperate to regulate autophagy (*22*). We further confirmed, by in vitro ubiquitylation assays, that UBA6 was also the preferred E1 for BIRC6-dependent ubiquitylation of casp-3, casp-7, and HTRA2 (fig. S1C). To remove confounding effects from protease activity, BIRC6 was incubated with casp-3 or casp-7 in presence of the caspase inhibitor Z-VAD-FMK, or the protease-inactivating mutant $Ser^{306}\rightarrow$Ala of HTRA2 (hereafter HTRA2-S306A). In presence of UBA6, ubiquitin, and adenosine triphosphate (ATP), efficient ubiquitylation of all three proteases was observed (Fig. 1D). This activity is strictly dependent on IBM binding through the BIR domain, as it is lost upon introduction of inactivating mutations (C328S/C331S, referred to as BIRmut) (*23*, *34*) (Fig. 1D). The ubiquitylation of all substrates is inhibited by SMAC (Fig. 1D), which itself is also a substrate for BIRC6 ubiquitylation in vitro (*23*, *25*) (fig. S1, D and E). A mutant form of SMAC (SMAC*, IBM motif mutated to MVPI), which no longer binds to BIR domains (*23*), fails to inhibit BIRC6 activity (Fig. 1D). In the absence of protease inhibitors, casp-3, casp-7, and HTRA2 digest BIRC6 in a strictly BIR domain–dependent fashion, which can be inhibited by SMAC but not SMAC* (Fig. 1E). Together, these findings establish that BIRC6 exerts the biochemical activities of an IAP using its BIR and UBC domains to bind and ubiquitylate casp-3, casp-7,

[1]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA. [2]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA.
*Corresponding author. Email: eric_fischer@dfci.harvard.edu

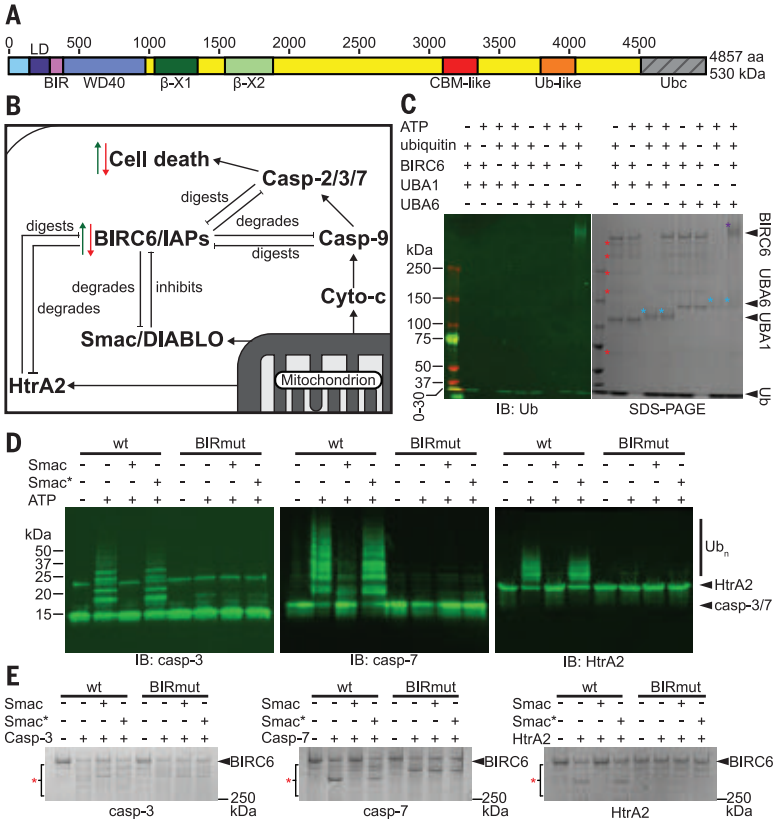


Fig. 1. BIRC6 is a UBA6-dependent ubiquitin ligase for caspases. (**A**) Domain scheme for BIRC6. Domain coloring and labeling are held constant throughout the figures. (**B**) Schematic overview of reported cellular interactions of BIRC6. (**C**) Auto-ubiquitylation assay. Anti-ubiquitin Western blot (left) and SDS–polyacrylamide gel electrophoresis (SDS-PAGE) (right) analysis of BIRC6 auto-ubiquitylation. A purple asterisk indicates auto-ubiquitylated BIRC6, and red asterisks mark BIRC6 degradation bands. Blue asterisks indicate ubiquitin-charged E1. (**D**) Ubiquitylation assay. Western blots using the indicated antibodies of ubiquitylation assays establishing casp-3, casp-7, and HTRA2 as in vitro substrates of BIRC6. A BIR domain mutant (BIRmut, C328S/C331S) is not capable of ubiquitylating substrates. Addition of wild-type processed SMAC (N-terminal AVPI) inhibits activity, but not addition of a SMAC variant (SMAC*, N-terminal MVPI). (**E**) SDS-PAGE analysis of BIRC6 stability assays upon incubation with casp-3, casp-7, and HTRA2. All three proteases digest wild-type (wt) BIRC6 (black arrowhead), and the effect can be reverted to baseline degradation by adding SMAC (but not SMAC*). Baseline degradation of the BIRmut variant is insensitive to the presence of SMAC or SMAC*. The height of the 250-kDa marker is indicated, and the area marked with a red asterisk indicates BIRC6 degradation bands. All gels and blots are representative of at least three independent replicates.

and HTRA2, and that BIRC6 itself is a substrate for these proteases. The absence of multiple BIR domains and linker regions previously shown to be important for efficient caspase binding (*35*) raises the question of how BIRC6 interacts with caspases and its regulators SMAC and HTRA2.

Cryo-EM structure reveals horseshoe-shaped dimeric architecture

To structurally characterize BIRC6, we collected cryo-EM data of full-length BIRC6. Initial two-dimensional (2D) classification revealed the presence of high-quality particles but also preferred orientations leading to highly anisotropic three-dimensional (3D) reconstructions (fig. S2). Omitting 2D classification and directly employing 3D classification allowed retention of particles representing rare views, which mitigated the preferred orientation problem. Several rounds of classification and local refinements lead to high-quality maps of BIRC6 refined to resolutions of 2.0 to 3.0 Å (Fig. 2A, figs. S2 and S3, and table S1). BIRC6 presents itself as a large (180 Å by 170 Å by 120 Å) head-to-tail dimer with an extended helical arch forming the dimer interface (~10,286 Å^2 interface area) and serving as the backbone for functional domains (Fig. 2B). The very N terminus forms a WD40-like propeller domain [amino acids (aa) 68 to 966)], which has a linker domain (LD; aa 151 to 265) and the BIR domain (aa 278 to 364) protruding out between individual blades of the propeller (Fig. 2C). The IBM binding groove on the BIR domain is facing the central cavity of BIRC6, as revealed by superposition (main-chain root mean square deviation: 1.2 Å) with a crystal structure of the X-linked inhibitor of apoptosis protein (XIAP) BIR3 domain in complex with the AVPI peptide of SMAC (*36*) (Fig. 2C and fig. S3H). The WD-40, LD, and BIR domains are connected to the helical arch by a disordered linker (~40 aa) and are sitting on two beta-sandwich domains [β-X1 (aa 1030 to 1323) and β-X2 (aa 1563 to 1884)] (Fig. 2B). Centrally inserted into the helical arch is a carbohydrate-binding module family 32 (CBM)–like domain (*37*) (CBM, aa 3160 to 3302) (Fig. 2, B and D), followed by an unpredicted ubiquitin-like domain (Ubl, aa 3819 to 4068) close to the C-terminal UBC domain (aa 4520 to 4857) (Fig. 2E). The UBC domain itself is invisible in the cryo-EM reconstructions owing to its high degree of flexibility, which likely enables efficient ubiquitylation of diverse substrates.

Tight binding of SMAC to BIRC6 facilitates a caspase release mechanism

Previous reports found similar affinities for full-length SMAC and caspase-IBM binding to individual BIR domains of XIAP and cellular inhibitor of apoptosis protein 1 (cIAP1) with binding affinity (K_d) values in the range of hundreds of nanomolar to micromolar (*38*–*40*). This would suggest that a large excess of SMAC is necessary for efficient release of bound caspases. Binding affinities, however, are greatly increased with larger constructs of IAPs containing multiple BIR domains (*39*), which led to the conclusion that caspases and SMAC compete for mutual binding sites through multivalent interactions spanning multiple domains (*39*, *41*–*43*). In combination with structural studies, a model with two distinct BIR domains tucked under an arch-shaped SMAC dimer was proposed (*36*, *39*, *43*) (fig. S4A). This model cannot apply to BIRC6, which has only a single BIR domain, and even if one BIR domain was contributed from each protomer, they are still too far apart from each other to form the proposed arrangement (fig. S4A). We thus set out to characterize the binding mode of SMAC to BIRC6 and established a time-resolved Förster resonance energy transfer (TR-FRET)–based equilibrium binding assay to determine binding constants of SMAC, casp-3, casp-7, and HTRA2-S306A. Substrates and BIRC6 were labeled with BODIPY and a terbium-pentafluorophenyl (Pfp) ester (*44*), respectively. To conduct equilibrium measurements, caspases were inhibited by addition of Z-VAD-FMK, which is not expected to change the IBM-driven binding to BIRC6. SMAC binding is the tightest observed, with an apparent K_d of <2 nM (assay limited), followed by casp-7 and HTRA2-S306A with K_d of

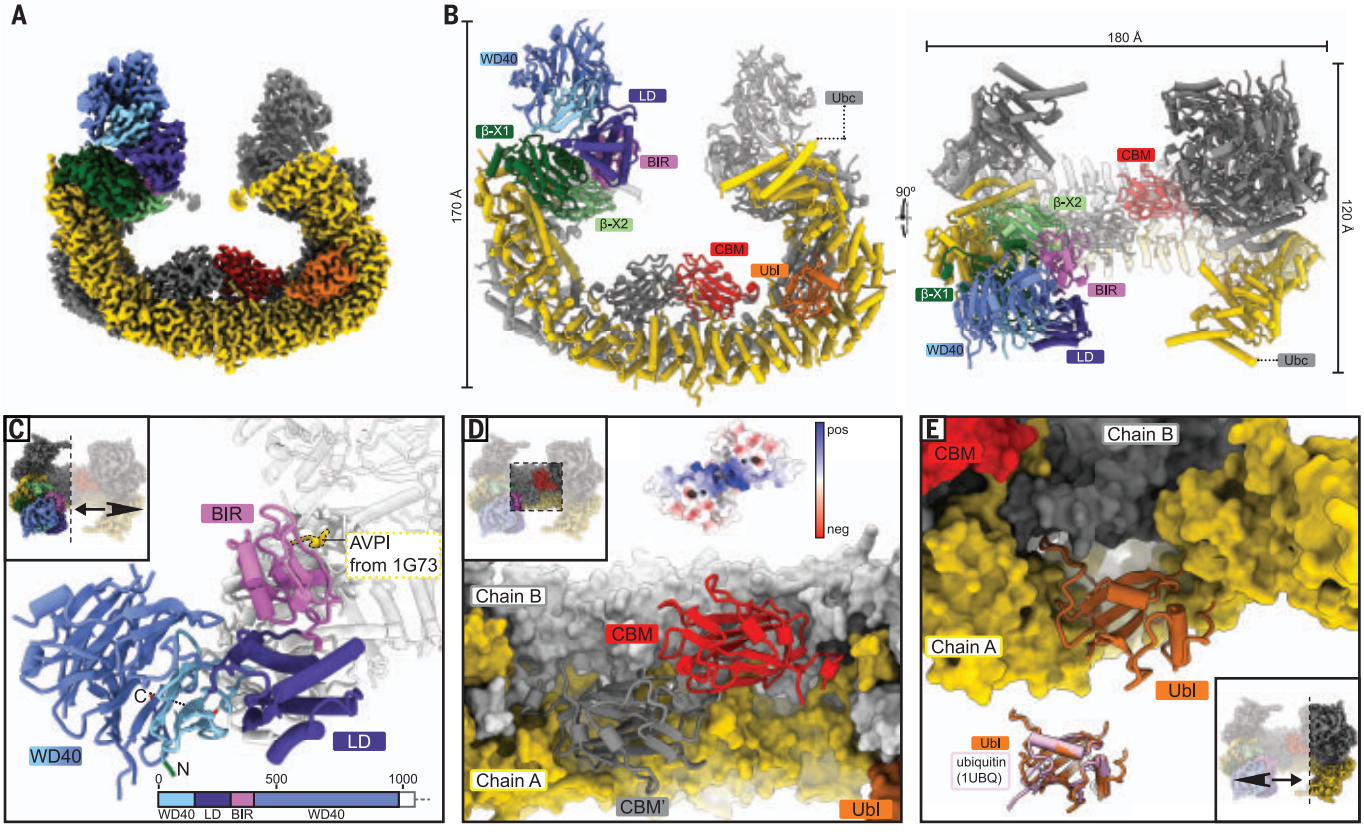


Fig. 2. Cryo-EM structure of full-length human BIRC6. (**A**) Composite EM map of the BIRC6 dimer, with one chain colored as in Fig. 1A and the other chain colored in gray. (**B**) Cartoon model of BIRC6 in two orientations. Domains of one protomer are labeled. The UBC domain is not visible in the reconstructions. (**C**) The N-terminal ~1000 amino acids comprise a disconnected WD40 propeller with the LD and BIR domains inserted between blades 2 and 3. The location of the peptide binding grove on the BIR domain is illustrated by placement of the AVPI peptide taken from PDB ID 1G73 (*36*). The N and C termini are colored green and red, respectively, and the disordered linker between the C terminus and the β-X1 domain is indicated by a dashed line behind the domains. (**D**) Close-up on the CBM-like domains, holding the dimer together in a ball clasp fashion. The surface of the two domains colored according to electrostatic potential is shown at the top, revealing a highly positively charged path right in the center. (**E**) Zoomed-in view of the unpredicted ubiquitin-like domain (Ubl). An overlay with ubiquitin [PDB ID 1UBQ (*54*)] is shown at the bottom.

4 ± 1 nM (assay limited) and 75 ± 5 nM, respectively (fig. S4, B and C). Binding of casp-3 was too weak to establish a K_d. To determine the rank order of competitive binding relevant for regulation, we established a TR-FRET displacement assay where BODIPY-labeled HTRA2-S306A was displaced from BIRC6 by titration of unlabeled SMAC, casp-3, casp-7, and HTRA2-S306A with half-maximal inhibitory concentration (IC_{50}) values of 9 ± 1, 167 ± 31, 383 ± 60, and 1358 ± 148 nM for SMAC, casp-7, HTRA2, and casp-3, respectively (Fig. 3A). These affinities are consistent with mature SMAC released from the mitochondria effectively inhibiting the anti-apoptotic activity of BIRC6 by blocking interactions with caspases.

To test whether SMAC binding is solely governed by interactions of the SMAC IBM motif and the BIR domains, we repeated the experiment with the BIRmut variant of BIRC6 (Fig. 3B). SMAC still bound with an apparent K_d of 26 ± 2 nM, confirming the existence of a previously proposed BIR-independent binding site for SMAC on BIRC6 (*26*). To characterize the binding mode, we collected cryo-EM data for a stable BIRC6-SMAC complex and obtained an anisotropic reconstruction at a nominal resolution of 2.5 Å (fig. S4D and table S2). The resulting maps revealed clear additional helical density in the central cavity (Fig. 3, C and D). Three-dimensional classification and local refinements enabled unambiguous placement of dimeric SMAC using a previously determined crystal structure (*45*) (aa 12 to 184) (Fig. 3, D and E). Because of limited map quality in this region, the SMAC N termini in the IBM binding grooves on the BIR domains were not resolved. A local resolution-filtered map, however, revealed a connection between the modeled SMAC N termini and the BIR domains (Fig. 3C), which is accounted for by unmodeled residues (Fig. 3E). Notably, we also identified clear symmetrical density for two additional helices on top of the SMAC dimer that are not from SMAC itself (Fig. 3, D and E). Although the quality of the map prevented unambiguous building of the helix, careful inspection of secondary structure predictions and unmodeled parts identified it as an extended helix contributed by BIRC6 (aa 1616 to 1666), connected to the β-X2 domain through long disordered linkers of ~34 and 105 aa. The two helices, one contributed from each protomer, hold SMAC in place on top of the CBM domains (Fig. 3E). To determine whether these helices constitute the BIR-independent secondary binding site (*26*), we measured the equilibrium binding constants of a deletion construct of BIRC6 (ΔHelix, aa 1616 to 1666 removed) to be 30 ± 5 nM, comparable to the BIRmut construct (Fig. 3B). A double mutant (BIRmut/ΔHelix) had no observable affinity to SMAC (Fig. 3B), consistent with BIR domain and helix constituting the two binding sites. In vitro ubiquitylation assays confirmed that there is a

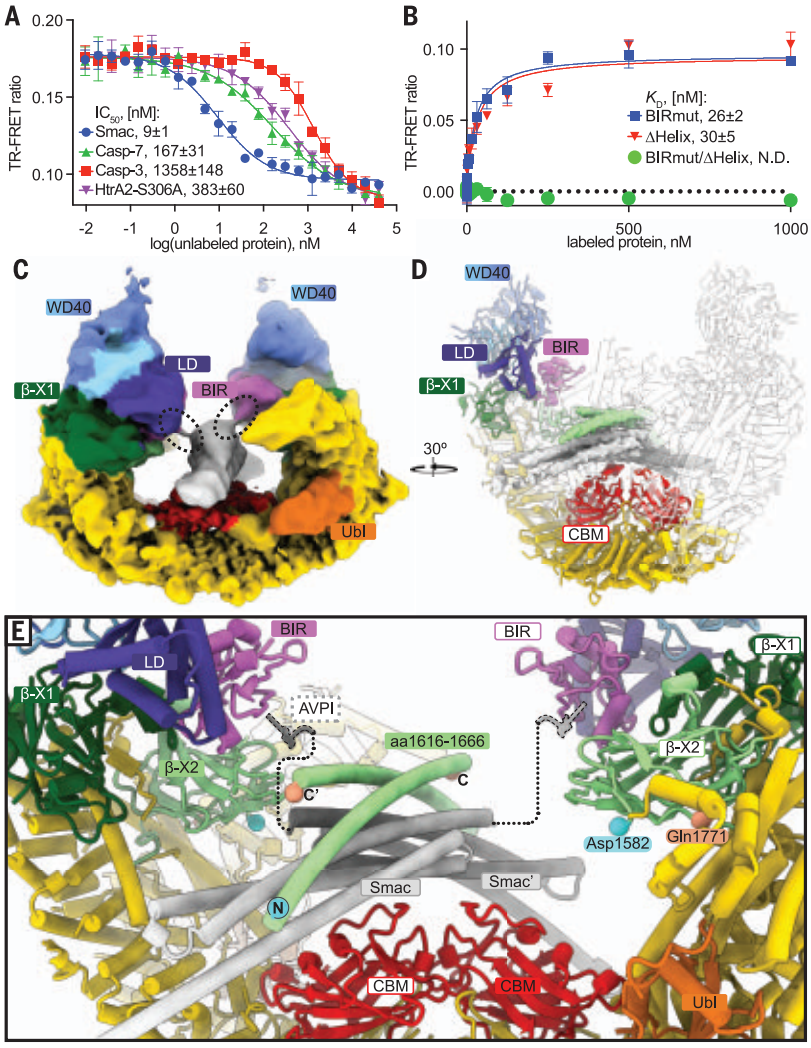


Fig. 3. Cryo-EM structure of BIRC6 in complex with SMAC. (**A**) TR-FRET–based displacement assay. Increasing concentrations of SMAC, casp-7, HTRA2-S306A, and casp-3 were titrated into BODIPY-labeled HTRA2-S306A at 50 nM, Flag-BIRC6 at 10 nM, and Tb-anti-Flag antibody at 8 nM. IC_{50} values of 9 ± 1, 167 ± 31, 383 ± 60, and 1358 ± 148 nM were determined for SMAC, casp-7, HTRA2-S306A, and casp-3, respectively. (**B**) TR-FRET equilibrium binding assay of a BIRC6 BIR-domain mutant, a ΔHelix variant, a double mutant, and BODIPY-labeled SMAC. The single mutants show comparable K_d of 26 ± 2 and 30 ± 5 nM for BIRmut and ΔHelix, respectively, whereas the double mutant has no detectable binding. Data in (A) and (B) are represented as mean ± SD from three technical replicates. N.D., not determined. (**C**) Local resolution-filtered map highlighting the connectivity between SMAC (gray) and the BIR domains of BIRC6 (in color), indicated by dashed circles. (**D**) Side view showing locally refined SMAC density in the central BIRC6 cavity. Density for the additional helices is colored light green like the β-X2 domain. (**E**) Close-up highlighting how the SMAC dimer is arching over the CBM domains. The connection to the IBM groove on the BIR domains is indicated with a dashed line, and the shown AVPI peptide is modeled after PDB ID 1G73 (*36*). The attachment points of the two additional helices (directionality indicated with labels for N- and C-terminal ends) with the β-X2 domain are indicated with matching spheres and residue numbering. Domain labels with solid backgrounds belong to one protomer, domain labels with white backgrounds indicate domains of the second protomer.

cumulative decrease of activity toward SMAC (fig. S5A), without apparent change in auto-ubiquitylation activity of the different mutants (fig. S5B). A study published in this same issue (*46*) observed similar additional density and assigned it to two helices (~aa 2222 to 2632 and ~aa 2268 to 2300) contributed from only one protomer. Although additional contacts with unresolved loop regions or alternative binding modes are conceivable, the proposed arrangement is incompatible with the density we observe and our mutational studies. In addition to the two described interactions, there is substantial charge complementarity between SMAC and the CBM domains (Fig. 2D and fig. S4H). While performing competitive binding assays, we noticed that SMAC, once bound to BIRC6, exhibited very slow off rates compared with HTRA2, with virtually no displacement visible up to 22 hours, whereas HTRA2 was readily displaced (fig. S5, C and D). This near-irreversible binding of dimeric SMAC is in accordance with the structural arrangement and explains how SMAC can expel bound caspases when released from mitochondria and thereby drive apoptosis. These findings are consistent with our data showing that BIRC6 ubiquitylation of casp-3 and casp-7 is effectively inhibited by the presence of SMAC (Fig. 1D).

Caspases bind to the BIR domain and reside in the central cavity

BIRC6 does not contain multiple BIR domains or the linker region that was found to be essential for XIAP binding to casp-3 and casp-7 (*35*, *42*). We thus set out to characterize the binding of these caspases to full-length BIRC6. To reconstitute a stable complex between BIRC6 and casp-3 or casp-7, we added Z-VAD-FMK to prevent proteolysis of BIRC6. Reconstructions from datasets collected for BIRC6 bound to casp-3 or casp-7 revealed a similar overall architecture for BIRC6 (fig. S6 and table S2). In both structures, the caspase occupies the central cavity established by the two arms of the arch and the two CBM domains in a flexible manner, which manifests in blurred density (Fig. 4A and fig. S6). Despite the lack of high-quality density for the caspase, we could place crystal structures of the corresponding dimeric casp-3 (*35*), constrained by the position of the N termini and the BIR domains (Fig. 4, B to E). Connecting density (Fig. 4A) between the caspase and the BIR domains together with our mutational data (Fig. 1E) confirms that the dimeric caspases are recruited and held in place by canonical BIR domain interactions with the caspase IBM (Fig. 4E). The overall negatively charged active site region of casp-3 (Fig. 4C), which in previous structures in complex with the XIAP BIR2 domain is occupied by the linker preceding BIR2 (*35*), is oriented toward the positively charged CBM domains (Fig. 4, D and E). In some clusters (clusters 2 and 3 in Fig. 4A), there appears to be a connection between the CBM and caspase density, which could potentially be accounted for by a positively charged loop of the CBM (aa 3186 to 3196) interacting with the negatively charged active site region (Fig. 4, C to E). This observation suggests that the BIRC6–caspase interaction is further stabilized, in addition to IBM binding by the BIR domains, by extended charge and shape complementarity (Fig. 4, C to E). This hypothesis has recently been substantiated by the fact that charge-reversal mutations in the CBM loop impair BIRC6-mediated inhibition of casp-3 and ubiquitylation of casp-7 (*31*).

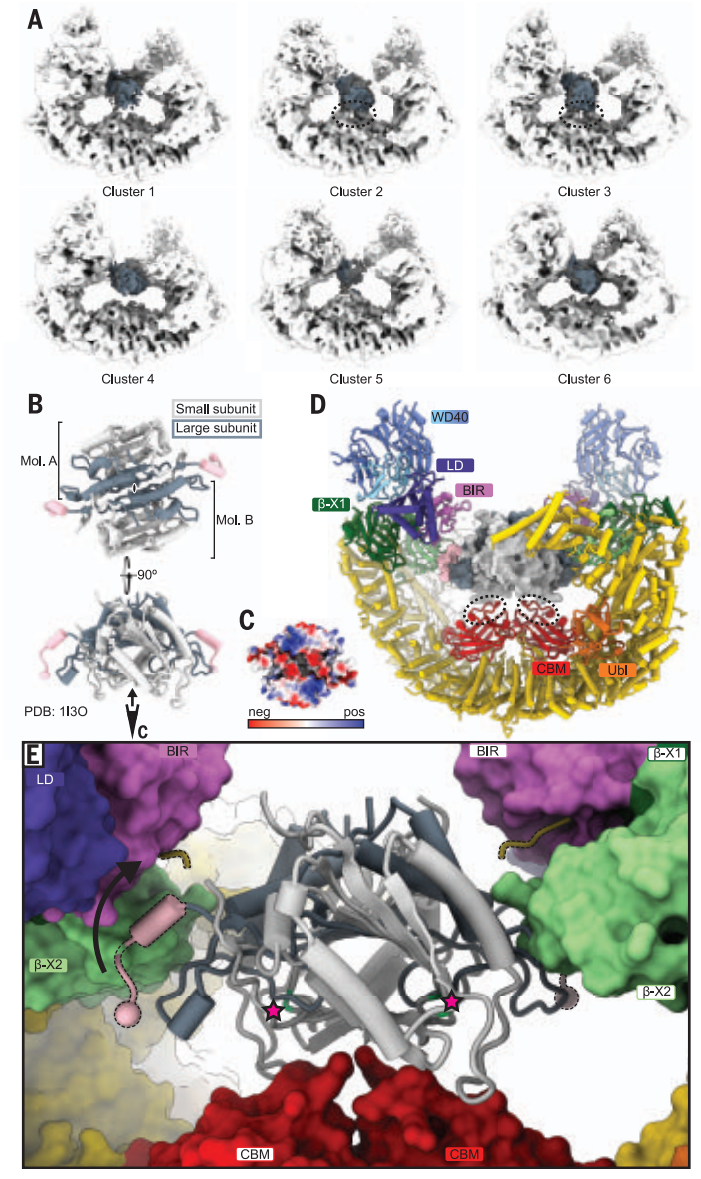


Fig. 4. Model for the BIRC6/casp-3 complex. (**A**) Six particle clusters after 3D variability analysis revealing a highly flexible binding mode of casp-3 in the central cavity of BIRC6. BIRC6 is shown in white, casp-3 in gray, and density connecting to the CBM domains in clusters 2 and 3 is indicated with dashed lines. (**B**) Casp-3 structure from a published casp-3/XIAP-BIR2 crystal structure [PDB ID 1I3O (*35*); BIR2 domain removed for clarity], with the large subunit in dark gray, the small subunit in light gray, and the IBM in salmon. Two orientations are shown, and the viewing direction for (C) is indicated. (**C**) Electrostatic potential map, viewed as indicated in (B), revealing a negatively charged patch near the caspase active sites. (**D**) Casp-3 shown as surface in same orientation as in (B), bottom, modeled in the central BIRC6 cavity indicating a snug fit. The positively charged loop (aa 3186 to 3196) in the CBM is indicated by dashed lines. (**E**) Zoomed-in view for the BIRC6/casp-3 model, with casp-3 shown as cartoon in same orientation as in (B), bottom, and (D). BIRC6 is shown as surface, and the location of the IBM groove on the BIR domain is shown by a dashed yellow cartoon strand taken from PDB ID 1I3O (*35*). The active-site dyad of casp-3 (H121/C163) is colored green and highlighted with pink stars. The IBM of casp-3 needs to swing up to reach into the BIR domain; this movement is indicated with a black arrow.

HTRA2 binds BIRC6 competitively with caspases and SMAC

Similar to SMAC, the serine protease HTRA2, which is primarily involved in the clearance of misfolded proteins in the intermembrane space (*47*), is also released from the mitochondria upon activation of the apoptotic cascade (*48*). HTRA2 is a trimeric enzyme and after maturation in the mitochondrion consists of an N-terminal protease domain and a C-terminal PDZ domain, which is used for autoregulation and substrate recruitment (*49*–*51*) (fig. S7A). When released from the mitochondria, a canonical IBM (AVPS, fig. S1A) is exposed, and HTRA2 has been shown to interact with BIRC6 (*23*). We sought to determine the structure of HTRA2 bound to BIRC6 to better understand how the trimeric HTRA2 is engaged by the dimeric BIRC6. A stable complex of BIRC6 and HTRA2-S306A was purified and its consensus structure determined by cryo-EM at an overall resolution of 2.8 Å (fig. S7, B to G, and table S2). Density was observed in the central cavity of BIRC6 and, after 3D variability analysis followed by clustering, identified as a trimer of HTRA2-S306A (Fig. 5A). Making extensive contacts to the CBM domains and the helical arch structure of BIRC6, HTRA2-S306A is perfectly placed to be anchored by the BIR domains (Fig. 5, B and C). The trimeric protease is offset from the BIRC6 dimer axis, resulting in an asymmetric assembly with one of the three protease domains centered over the positively charged loops of the CBM domains (fig. S7F). Overall, the assembly shows high plasticity (fig. S7B), with all three IBM motifs on HTRA2 in proximity to the BIR domains. HTRA2-S306A presents itself in a previously unobserved conformation that appears to be primed for activity (fig. S7A), with only one of the PDZ domains clearly defined by density (Fig. 5A). This intermediate conformation is stabilized by the Ubl domain in BIRC6 (Fig. 5, B and C). Similar to SMAC and caspases, HTRA2 interacts with the CBM domains primarily through an extended interface of charge complementarity (fig. S7F), and the overall conformation suggests that HTRA2, upon engagement with BIRC6, enters into a conformation that is no longer autoinhibited, explaining the observed proteolysis of BIRC6 by HTRA2 (Fig. 1E). In the observed arrangement, the active sites of HTRA2 are only ~24 Å away from a potential HTRA2 recognition site (SYIF) (*52*), which is located in a loop in the CBM domains (aa 3198 to 3201) (Fig. 5C), and the observed flexibility could well allow HTRA2 to cut BIRC6 at this site.

SMAC mimetics have a modest effect on BIRC6 activity

Given that BIRC6 is a negative regulator of apoptosis in cells (*23*) and exhibits all the canonical IAP biochemical features, we asked whether SMAC mimetics designed to promote apoptosis in tumors by blocking caspase recruitment to IAPs would also inhibit BIRC6 activity. We obtained representatives of all classes of clinical stage SMAC mimetics: SM-406, birinapant, GDC-0917, LCL-161, GDC-0152, and ASTX660 (fig. S8). To establish potency in vitro, we performed casp-3 ubiquitylation assays at increasing concentrations of each inhibitor (Fig. 6A). Overall, we observe modest inhibitory activity, with a rank order of LCL-161, GDC-0152, GDC-0917, SM-406 > ASTX660 >

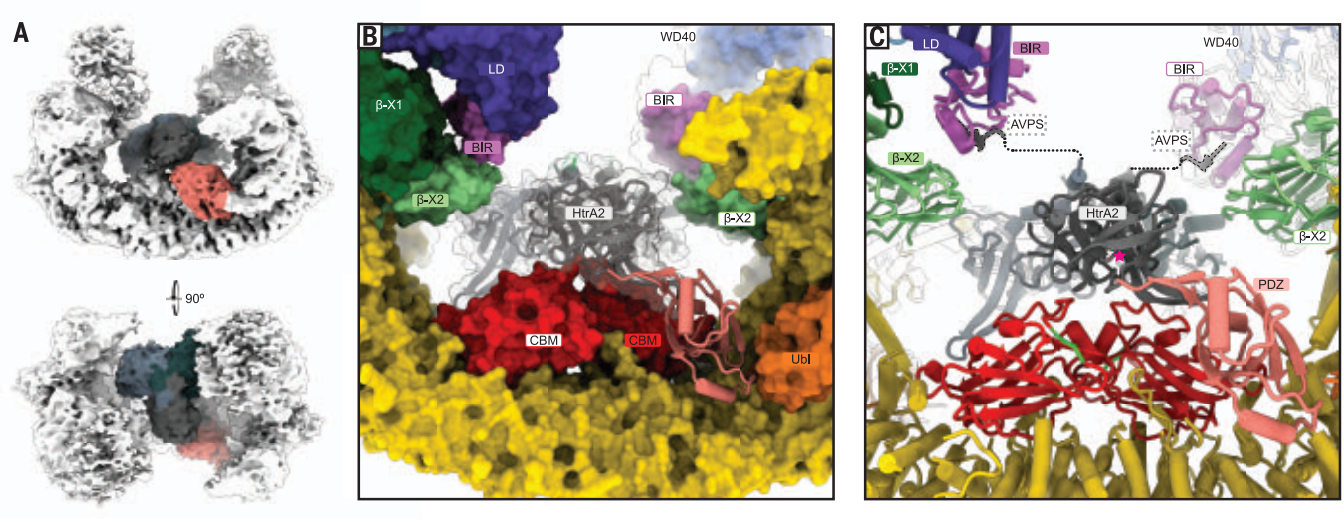


Fig. 5. Cryo-EM structure of BIRC6 in complex with HTRA2. (**A**) Cryo-EM density, low-pass filtered to 6 Å, of BIRC6 (white) in complex with HTRA2-S306A (colored). (**B**) Detailed view with BIRC6 shown as a surface representation and HTRA2-S306A shown as a cartoon. The PDZ domain (salmon) wedges in between the Ubl and CBM domains, while the protease domains sit over the CBM domains. (**C**) Detailed view with BIRC6 and HTRA2-S306A shown as cartoons. The location of the HTRA2-S306A active site closest to BIRC6 is indicated with a pink star, and the putative protease recognition site on BIRC6 (in the CBM) is highlighted in green.

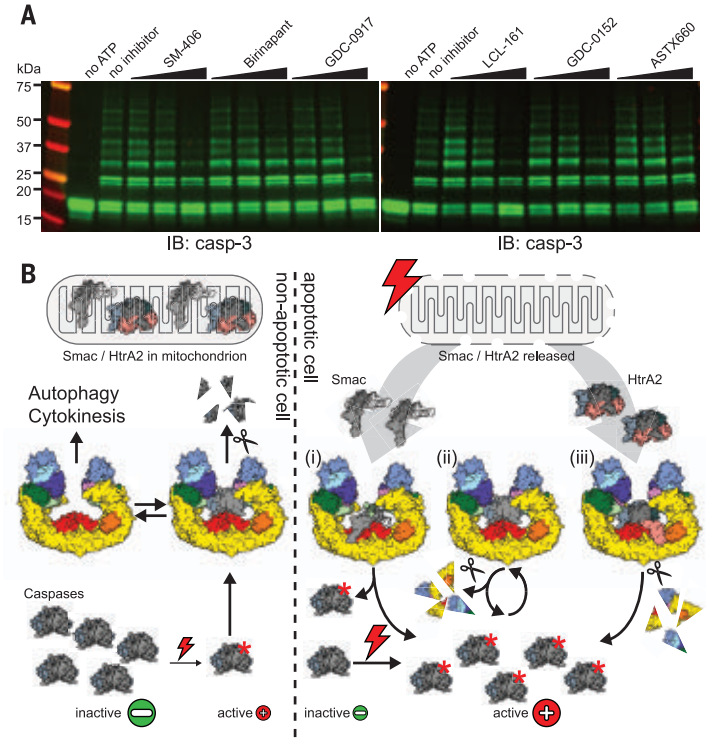


Fig. 6. SMAC mimetics have a limited effect on BIRC6 activity. (**A**) In vitro ubiquitylation assays. In vitro ubiquitylation of casp-3 by BIRC6 in the presence of increasing concentrations (1, 10, and 50 μM) of the indicated inhibitors was visualized by anti–casp-3 Western blots. Blots are representative of three independent replicates. (**B**) Cartoon summary illustrating the different regulatory networks in non-apoptotic (left) and apoptotic (right) cells. In non-apoptotic cells, caspases are largely inactive, and the small amount of active caspase is immediately cleared from the cell through proteasomal degradation. BIRC6 is able to regulate other pathways at the same time. Upon trigger of apoptosis, SMAC and HTRA2 are released from the mitochondrion, and the amount of active caspase is increased, leading to (i) inhibition of BIRC6-caspase binding by SMAC and (ii) proteolytic cleavage of BIRC6 by active caspases. HTRA2, while not able to efficiently free bound caspases, is also cleaving BIRC6 (iii), and all these mechanisms combined lead to a cumulative increase of active caspases.

birinapant. These findings are in line with the fact that these SMAC mimetics were optimized for binding to cIAPs and XIAP (*15*) and exhibit weak affinity for the BIR domain of BIRC6. It is unlikely that bivalent SMAC mimetics, such as birinapant, can engage both BIR domains of BIRC6 simultaneously given the distance between the two. Hence, they will not exhibit improved efficacy, in line with the observed inability to effectively inhibit BIRC6-mediated ubiquitylation of casp-3.

Discussion

Our structural and biochemical characterization of BIRC6 and its interaction with key pro-apoptotic factors provides a molecular basis for BIRC6 activity as an IAP and an example of a full-length IAP protein engaging its clients (Fig. 6B). The dimeric architecture of this single BIR domain IAP is necessary for effective interaction with its multimeric substrates and explains why BIRC6 homodimerization is required for its IAP function (*23*). Only in this dimeric arrangement can two BIR domains simultaneously engage caspases, positioning all observed binding partners in the central cavity of BIRC6 in a highly mobile fashion (movie S1). The structures reveal considerable differences in caspase binding to what had been observed for other IAPs. In the case of XIAP, casp-3 and casp-7 were found to interact not only with the BIR2 domain itself but also with a linker region N-terminal to it (*35*, *42*, *53*). In our structure of BIRC6, casp-3 and casp-7 engage in a canonical BIR domain interaction with the active sites facing the CBM domains. In addition, the interaction is stabilized by electrostatic complementarity with

the CBM domains, confining casp-3 and casp-7 to the central cavity of BIRC6. This binding mode, together with the observation that SMAC binding is nearly two orders of magnitude tighter, with very slow off rates, explains how SMAC binding is mutually exclusive and can thereby effectively release active caspases from BIRC6 (Fig. 6B). The different affinities observed for casp-7, HTRA2, and casp-3 are at least partially encoded in their IBM (fig. S1A), with additional contributions from shape and charge complementarity. Casp-3, the weakest binder in our assays, presents the most distinct IBM, lacking the AXPX motif commonly found in type III BIR binding proteins (*9*). Although additional studies are needed to determine whether similar principles may apply to other IAPs, our structures now provide a molecular understanding for SMAC-mediated release of casp-3 and casp-7 from an intact IAP.

The mechanism observed for inhibition of caspases by BIRC6 and how it is counteracted by SMAC provides insights into two distinct ways of inhibiting highly processive enzymes such as proteases or E3 ubiquitin ligases. Although ubiquitylation and very tight, slow off-rate binding appear different at first, they both solve the problem of overcoming processivity through a kinetic component. Ubiquitin-mediated turnover leads to irreversible destruction of caspases, and very tight binding of SMAC to BIRC6 similarly leads to a near-irreversible sequestration of BIRC6, allowing apoptosis to proceed beyond the point of no return. This elegant solution is likely more general in the regulation of such enzymes. It remains to be seen how the tug-of-war between the proteolytic activity of the caspases and the ubiquitylation activity of BIRC6 plays out in a cellular context. It is conceivable that the scale is tipped toward one way or the other by a variety of different factors, including the relative abundance of all players involved. As noted before (*23*, *26*), it is also possible that one role of BIRC6 is to remove spuriously activated caspases under baseline conditions, whereas it can be overwhelmed upon induction of apoptosis.

Despite several generations of SMAC mimetics with differing profiles in selectivity and binding modes, the clinical responses have been underwhelming (*16*). All current SMAC mimetics were developed with a focus on XIAP and cIAPs, as the role of BIRC6 and other family members as anti-apoptotic IAPs remained ambiguous. Our data demonstrate that clinically explored SMAC mimetics are poor antagonists of BIRC6-mediated caspase degradation and thereby spare this arm of IAP-mediated prosurvival signals. Our structures and biochemical characterization now offer ways to directly target BIRC6. This could be achieved through bivalent SMAC-mimetic inhibitors with sufficient linker length to engage both BIR domains simultaneously, or by targeting the CBM domain of BIRC6 to block engagement with caspases.

REFERENCES AND NOTES

1. L. Galluzzi *et al*., *Cell Death Differ.* **25**, 486–541 (2018).
2. S. Elmore, *Toxicol. Pathol.* **35**, 495–516 (2007).
3. B. Favaloro, N. Allocati, V. Graziano, C. Di Ilio, V. De Laurenzi, *Aging* **4**, 330–349 (2012).
4. O. Julien, J. A. Wells, *Cell Death Differ.* **24**, 1380–1389 (2017).
5. J. Silke, P. Meier, *Cold Spring Harb. Perspect. Biol.* **5**, a008730 (2013).
6. E. C. LaCasse *et al*., *Oncogene* **27**, 6252–6275 (2008).
7. M. Hrdinka, M. Yabal, *Genes Immun.* **20**, 641–650 (2019).
8. M. Saleem *et al*., *Chem. Biol. Drug Des.* **82**, 243–251 (2013).
9. B. P. Eckelman, M. Drag, S. J. Snipas, G. S. Salvesen, *Cell Death Differ.* **15**, 920–928 (2008).
10. R. Singh, A. Letai, K. Sarosiek, *Nat. Rev. Mol. Cell Biol.* **20**, 175–193 (2019).
11. C. Candé *et al*., *Biochimie* **84**, 215–222 (2002).
12. A. M. Verhagen *et al*., *Cell* **102**, 43–53 (2000).
13. C. Du, M. Fang, Y. Li, L. Li, X. Wang, *Cell* **102**, 33–42 (2000).
14. A. M. Verhagen *et al*., *J. Biol. Chem.* **277**, 445–454 (2002).
15. L. Bai, D. C. Smith, S. Wang, *Pharmacol. Ther.* **144**, 82–95 (2014).
16. E. Morrish, G. Brumatti, J. Silke, *Cells* **9**, 406 (2020).
17. Z. Chen *et al*., *Biochem. Biophys. Res. Commun.* **264**, 847–854 (1999).
18. H. P. Hauser, M. Bardroff, G. Pyrowolakis, S. Jentsch, *J. Cell Biol.* **141**, 1415–1422 (1998).
19. C. Pohl, S. Jentsch, *Cell* **132**, 832–845 (2008).
20. C. Pohl, S. Jentsch, *Nat. Cell Biol.* **11**, 65–70 (2009).
21. R. Jia, J. S. Bonifacino, *Autophagy* **16**, 382–384 (2020).
22. R. Jia, J. S. Bonifacino, *eLife* **8**, e50034 (2019).
23. T. Bartke, C. Pohl, G. Pyrowolakis, S. Jentsch, *Mol. Cell* **14**, 801–811 (2004).
24. X. B. Qiu, S. L. Markant, J. Yuan, A. L. Goldberg, *EMBO J.* **23**, 800–810 (2004).
25. Y. Hao *et al*., *Nat. Cell Biol.* **6**, 849–860 (2004).
26. X. B. Qiu, A. L. Goldberg, *J. Biol. Chem.* **280**, 174–182 (2005).
27. C. Domingues, H. D. Ryoo, *Cell Death Differ.* **19**, 470–477 (2012).
28. S. Y. Vernooy *et al*., *Curr. Biol.* **12**, 1164–1168 (2002).
29. J. Silke, A. M. Verhagen, P. G. Ekert, D. L. Vaux, *Cell Death Differ.* **7**, 1275 (2000).
30. L. Cao, Z. Wang, X. Yang, L. Xie, L. Yu, *FEBS Lett.* **582**, 3817–3822 (2008).
31. L. Dietz *et al*., *Science* **379**, 1112–1117 (2023).
32. K. Sekine *et al*., *Biochem. Biophys. Res. Commun.* **330**, 279–285 (2005).
33. A. Tsherniak *et al*., *Cell* **170**, 564–576.e16 (2017).
34. F. Li *et al*., *Nature* **396**, 580–584 (1998).
35. S. J. Riedl *et al*., *Cell* **104**, 791–800 (2001).
36. G. Wu *et al*., *Nature* **408**, 1008–1012 (2000).
37. S. Shinya *et al*., *Biochem. J.* **473**, 1085–1095 (2016).
38. R. Kulathila *et al*., *Acta Crystallogr. D Biol. Crystallogr.* **65**, 58–66 (2009).
39. Y. Huang, R. L. Rich, D. G. Myszka, H. Wu, *J. Biol. Chem.* **278**, 49517–49522 (2003).
40. Z. Liu *et al*., *Nature* **408**, 1004–1008 (2000).
41. S. M. Srinivasula *et al*., *Nature* **410**, 112–116 (2001).
42. J. Chai *et al*., *Cell* **104**, 769–780 (2001).
43. Z. Gao *et al*., *J. Biol. Chem.* **282**, 30718–30727 (2007).
44. N. C. Payne, A. S. Kalyakina, K. Singh, M. A. Tye, R. Mazitschek, *Nat. Chem. Biol.* **17**, 1168–1177 (2021).
45. J. Chai *et al*., *Nature* **406**, 855–862 (2000).
46. J. F. Ehrmann *et al*., *Science* **379**, 1117–1123 (2023).
47. L. Vande Walle, M. Lamkanfi, P. Vandenabeele, *Cell Death Differ.* **15**, 453–460 (2008).
48. H. R. Liu *et al*., *Circulation* **111**, 90–96 (2005).
49. W. Li *et al*., *Nat. Struct. Biol.* **9**, 436–441 (2002).
50. N. N. MohamedMohaideen *et al*., *Biochemistry* **47**, 6092–6102 (2008).
51. M. Merski *et al*., *Cell Death Dis.* **8**, e3119 (2017).
52. L. M. Martins *et al*., *J. Biol. Chem.* **278**, 49417–49427 (2003).
53. F. L. Scott *et al*., *EMBO J.* **24**, 645–655 (2005).
54. S. Vijay-Kumar, C. E. Bugg, W. J. Cook, *J. Mol. Biol.* **194**, 531–544 (1987).

ACKNOWLEDGMENTS

We thank the staff at the Harvard Cryo-EM Center for Structural Biology for their outstanding support during grid screening and data collection. We acknowledge the SBGrid Consortium for assistance with software and high-performance computing. We thank N. C. Payne and R. Mazitschek for providing the terbium-Pfp ester NCP311-Tb. We also thank E. Bennett, M. Eck, N. Thomä, W. Fairbrother, and members of the Fischer lab for valuable input and critical feedback on the manuscript. **Funding:** Funding was provided by National Institutes of Health grant NCI P01CA066996 (E.S.F.) and Mark Foundation Emerging Leader Award 19-001-ELA (E.S.F.). **Author contributions:** M.H. and E.S.F. conceived of the study and designed the research plan. M.H. cloned and purified proteins and conducted all biochemical assays and cryo-EM structure determination. C.Y.J. performed activity assays with commercial inhibitors. M.H. and E.S.F. designed the experiments, and all authors analyzed and interpreted data. E.S.F. supervised the study and acquired funding. M.H. prepared figures, and M.H. and E.S.F. wrote the manuscript. All authors approved the final version of the manuscript. **Competing interests:** E.S.F. is a founder, scientific advisory board (SAB) member, and equity holder of Civetta Therapeutics, Lighthorse Therapeutics, Proximity Therapeutics, and Neomorph, Inc. (also member of the board of directors). E.S.F. is an equity holder and SAB member of Avilar Therapeutics and Photys Therapeutics and a consultant to Novartis, Sanofi, EcoR1 Capital, and Deerfield. The Fischer lab receives or has received research funding from Novartis, Ajax, Voronoi, Interline, Deerfield, and Astellas. **Data and materials availability:** Cryo-EM maps and coordinates have been deposited in the Electron Microscopy Data Bank (EMDB) and Protein Data Bank (PDB), respectively, under accession codes EMD-27832 (BIRC6 consensus, PDB ID 8E2D), EMD-27833 (BIRC6 helical arch, PDB ID 8E2E), EMD-27834 (BIRC6 N-terminal arm, PDB ID 8E2F), EMD-27835 [BIRC6 N-terminal arm (aa 68 to 966), PDB ID 8E2G], EMD-27836 (BIRC6 C-terminal arm, PDB ID 8E2H), EMD-27837 (BIRC6/SMAC full, PDB ID 8E2I), EMD-27838 (BIRC6/SMAC local refine, PDB ID 8E2J), EMD-27839 (BIRC6/casp-3 with clusters), EMD-27840 (BIRC6/casp-7 with clusters), and EMD-27841 (BIRC6/HTRA2-S306A, PDB ID 8E2K). Uncropped gels and Western blot source data are available in figs. S9 and S10.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade5750
Materials and Methods
Figs. S1 to S10
Tables S1 to S4
References (*55–80*)
MDAR Reproducibility Checklist
Movie S1

Submitted 24 August 2022; accepted 30 January 2023
Published online 9 February 2023
10.1126/science.ade5750

STRUCTURAL BIOLOGY

Structural basis for SMAC-mediated antagonism of caspase inhibition by the giant ubiquitin ligase BIRC6

Larissa Dietz[1], Cara J. Ellison[1], Carlos Riechmann[1], C. Keith Cassidy[1], F. Daniel Felfoldi[1]†, Adán Pinto-Fernández[2,3], Benedikt M. Kessler[2,3], Paul R. Elliott[1]*

Certain inhibitor of apoptosis (IAP) family members are sentinel proteins that prevent untimely cell death by inhibiting caspases. Antagonists, including second mitochondria-derived activator of caspases (SMAC), regulate IAPs and drive cell death. Baculoviral IAP repeat-containing protein 6 (BIRC6), a giant IAP with dual E2 and E3 ubiquitin ligase activity, regulates programmed cell death through unknown mechanisms. We show that BIRC6 directly restricts executioner caspase-3 and -7 and ubiquitinates caspase-3, -7, and -9, working exclusively with noncanonical E1, UBA6. Notably, we show that SMAC suppresses both mechanisms. Cryo–electron microscopy structures of BIRC6 alone and in complex with SMAC reveal that BIRC6 is an antiparallel dimer juxtaposing the substrate-binding module against the catalytic domain. Furthermore, we discover that SMAC multisite binding to BIRC6 results in a subnanomolar affinity interaction, enabling SMAC to competitively displace caspases, thus antagonizing BIRC6 anticaspase function.

Programmed cell death, such as apoptosis, is triggered by internal or external signals, ultimately activating caspases—a family of proteases. Although cell death is required for normal development, it must be tightly controlled. Several mechanisms exist to prevent cell death in the absence of signaling cues and to ensure that caspase activity is stringently regulated (*1*). First, caspases are expressed as inactive zymogens (procaspases) requiring proteolytic processing for activation, and second, a subset of inhibitor of apoptosis (IAP) proteins restrict caspase activity.

IAPs contain a signature baculoviral IAP repeat (BIR) domain that binds caspases among other substrates and, frequently, a C-terminal ubiquitin ligase domain responsible for attaching ubiquitin posttranslationally to target proteins. Of the eight mammalian IAPs, only three are well-characterized inhibitors of apoptosis: cellular IAP1 [cIAP1; baculoviral IAP repeat-containing protein 2 (BIRC2)], cIAP2 (BIRC3), and X chromosome–linked IAP (XIAP; BIRC4) (*2*). XIAP directly inhibits activated caspase-3, -7, and -9 through a conserved Asp-containing pocket in the BIR2 and BIR3 domains binding to the amino terminus of the processed caspase small subunit (*3*). cIAP1 and cIAP2 (cIAP1/2) do not physically restrict caspase activity but promote caspase degradation through ubiquitination and promote cell survival through the production of proinflammatory ubiquitin chains in inflammatory signaling pathways (*4*–*6*).

BIRC6 (also known as Apollon or BRUCE, the latter in mice) is a giant IAP (4857 amino acids) conserved from *Drosophila* to humans and contains the hallmark BIR domain and a ubiquitin conjugation domain (UBC) conferring E2 ubiquitin ligase activity. Genetic studies highlight an essential role of BIRC6 in development and apoptosis inhibition—*BRUCE* knockout mice are embryonically lethal as a result of placental defects or increased apoptosis levels (*7*, *8*). Further, cellular studies have detected dual E2 and E3 ubiquitin ligase activity [only identified in one other ligase (*9*)] and inhibition of caspase-3, -7, and -9 (*10*–*13*). Roles of BIRC6 in other cellular processes, such as autophagy, are also emerging (*14*–*17*).

IAPs themselves must be inhibited to drive apoptosis, achieved in the intrinsic pathway by two mitochondrial-released proteins: HtrA2 cleaves XIAP and cIAP1/2 (*18*, *19*), whereas second mitochondria-derived activator of caspases (SMAC) directly competes with caspases binding to XIAP (*20*, *21*). BIRC6-mediated inhibition of apoptosis has been shown to be suppressed by SMAC in cellular studies (*10*, *22*) through uncharacterized mechanisms. Despite this clear cellular and genetic evidence for BIRC6 playing a critical role in inhibiting cell death, little is known about BIRC6 ubiquitin ligase activity, antiapoptotic mechanisms, and antagonism by SMAC.

BIRC6 functions exclusively with UBA6

To identify BIRC6 regulatory proteins, we expressed and purified recombinant full-length BIRC6 (fig. S1A) and performed mass spectrometry on affinity-purified complexes from unstimulated HEK293F cells. This approach revealed significant enrichment of the noncanonical E1, UBA6 (Fig. 1A and fig. S1B). Two E1 ubiquitin ligases exist with UBA1 responsible for initiating most ubiquitination cascades. UBA6, however, functions with a small subset of E2 enzymes and facilitates the transfer of the ubiquitin-like protein FAT10 in addition to ubiquitin (*23*, *24*). Notably, BIRC6 only receives ubiquitin from UBA6 and not UBA1 in in vitro transthiolation reactions (Fig. 1B and fig. S1C), consistent with earlier studies defining a UBA6-mediated role of BIRC6 in autophagy (*16*). BIRC6 weakly accepts FAT10 from UBA6 compared with UBE2Z, the known FAT10 acceptor (*25*) (fig. S1D). Our findings place BIRC6 as the only UBA6-specific, ubiquitin-selective ligase (fig. S1E).

Next, we investigated whether BIRC6 E2 activity is restricted to specific E3 families. In in vitro autoubiquitination assays testing a panel of E3s representative of the three families, BIRC6 displays cross-family E2 activity working with UBE3C and ARIH1 of the HECT and RBR E3 families, respectively (fig. S1, F to H). With more than 600 known E3 ubiquitin ligases, it is likely that BIRC6 also functions as a stand-alone E2 to additional E3 ubiquitin ligases.

BIRC6 restricts caspase activity

We also identified strong enrichment of the intrinsic apoptosis antagonists SMAC and HtrA2 in complex with BIRC6 (Fig. 1A), which led us to test BIRC6 E2-E3 activity in ubiquitinating critical components of intrinsic apoptosis. We observed robust ubiquitination of the activated intrinsic initiator caspase, caspase-9, and the activated executioner caspase-3 and -7 with weak ubiquitination of procaspases (Fig. 1C and fig. S2A). Additionally, BIRC6 ubiquitinates SMAC and inactive HtrA2 (S306A) (Fig. 1D and fig. S2B) but does not ubiquitinate activated caspase-8, the initiator caspase of extrinsic apoptosis (Fig. 1C and fig. S2C). In all instances, BIRC6 catalyzes multimonoubiquitination of substrates (fig. S2D) in contrast to polyubiquitination by XIAP and cIAP1 (fig. S2, E to H). Despite weak FAT10 loading onto BIRC6, we did not detect any transfer of FAT10 onto these substrates (fig. S1I). Our results show that BIRC6 functions as a combined E2 and E3 ubiquitin ligase, ubiquitinating key players in intrinsic apoptosis.

Because we observed robust ubiquitination of activated caspases by BIRC6, we tested whether BIRC6 can directly inhibit caspase activity. We recorded tight, nanomolar association between BIRC6 and activated caspase-9, -3, and -7 [dissociation constants (K_d) of 54.3 ± 8.8 nM, 8.9 ± 2.2 nM, and 10.3 ± 6.3 nM, respectively] (Fig. 1E). The addition of stoichiometric excess of BIRC6 impaired activated caspase-9 activity but only in the absence of the apoptosome complex (fig. S2, I to L), the macromolecular complex responsible for caspase-9 activation (*26*). By contrast, we observed that

[1]Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, UK. [2]Target Discovery Institute, Centre for Medicines Discovery, University of Oxford, Oxford, OX3 7FZ, UK. [3]Chinese Academy for Medical Sciences Oxford Institute, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7FZ, UK.
*Corresponding author. Email: paul.elliott@bioch.ox.ac.uk
†Present address: Department of Physics and Astronomy, University College London, London, WC1E 6BT, UK.

BIRC6 directly inhibits both caspase-3 and -7 activities in fluorogenic cleavage assays (Fig. 1, F and G). Our data reveal dual anticaspase activity of BIRC6: BIRC6 directly inhibits activated executioner caspase-3 and -7 and indirectly inhibits activated caspases by multimonoubiquitination, leading to caspase degradation in a cellular context.

BIRC6 architecture defines antagonism by SMAC

We next explored whether SMAC regulates the antiapoptotic activity of BIRC6. In a competition fluorogenic cleavage assay, we found that BIRC6-mediated caspase-3 inhibition reduces as the concentration of SMAC increases (Fig. 1H), showing that SMAC antagonizes BIRC6-caspase inhibition. To reveal how antagonism is achieved, we used single-particle cryo–electron microscopy (cryo-EM) analysis to determine the structures of BIRC6 alone or in complex with SMAC (Fig. 2, A to C; figs. S3 and S4; and table S1). Our structures reveal that BIRC6 has an overall crab-like architecture made up of an antiparallel dimer, consistent with size exclusion chromatography–multiangle light scattering (SEC-MALS) (fig. S3E), consisting of a rigid α-solenoid core (residues 1010 to 4502) (resolved to 2.9 Å) (fig. S6, A and B) with a flexible N-terminal substrate-binding module and an unresolved, highly flexible C-terminal catalytic domain acting as antennae and claws, respectively, guiding potential substrates toward a central, mouth-like cavity (Fig. 2, B and C). The antiparallel arrangement juxtaposes the substrate-binding module with the catalytic ubiquitin ligase domain at each end, potentially allowing for both intra- and interchain activity.

Our structures reveal multiple, previously unannotated domains interwoven around the solenoid core, contributing to BIRC6 architecture and function (Fig. 2A). In the central cavity, we identify a DOC domain (residues 3160 to 3308), an all-beta strand structure first identified for substrate binding in the anaphase-promoting complex/cyclosome (APC/C) APC10 subunit (*27*). APC10 functions as a substrate-binding module when in combination with coactivators Cdc20 or Cdh1, recruiting D-box containing substrates into proximity of the APC/C catalytic module (*28*). DOC domains have since been identified in other E3 ubiquitin ligases (*29*) (fig. S5, A to C), where they also likely serve as substrate-binding platforms.

We observe two key roles for the central DOC domain in BIRC6. First, a structural role in which a short loop from the core of one chain snakes over and nestles within the solenoid core of the adjacent chain, positioning the DOC domain on top of the other polypeptide chain and interlocking the two chains (Fig. 2D). The DOC domain is further stabilized by a beta strand (residues 2875 to 2889) protruding ~250 residues before the DOC domain (colored red in Fig. 2D and fig. S5C). Second, we determine a key functional role of the DOC domain—in the BIRC6-SMAC complex, we observed density (~8 Å) above this central DOC domain platform, allowing us to place the dimeric structure of SMAC in this pivotal position (Fig. 2, C and E, and fig. S7E). In contrast to other DOC domain proteins, BIRC6 has a loop extending from the DOC domain with a positively charged tip (residues 3189 to 3193; HRRAR) (Fig. 2D and fig. S5, A and C). The antiparallel dimeric nature of BIRC6 positions each HRRAR loop extension toward each other, forming an extended substrate-binding interface. In BIRC6, we observe that each loop makes contact with the SMAC helical bundle (Fig. 3C and fig. S5C), which indicates that this distinctive extension is important for substrate engagement.

In the N-terminal arm, we identify a DOC-like domain and a jelly-roll domain, both with potential for substrate binding and positioning

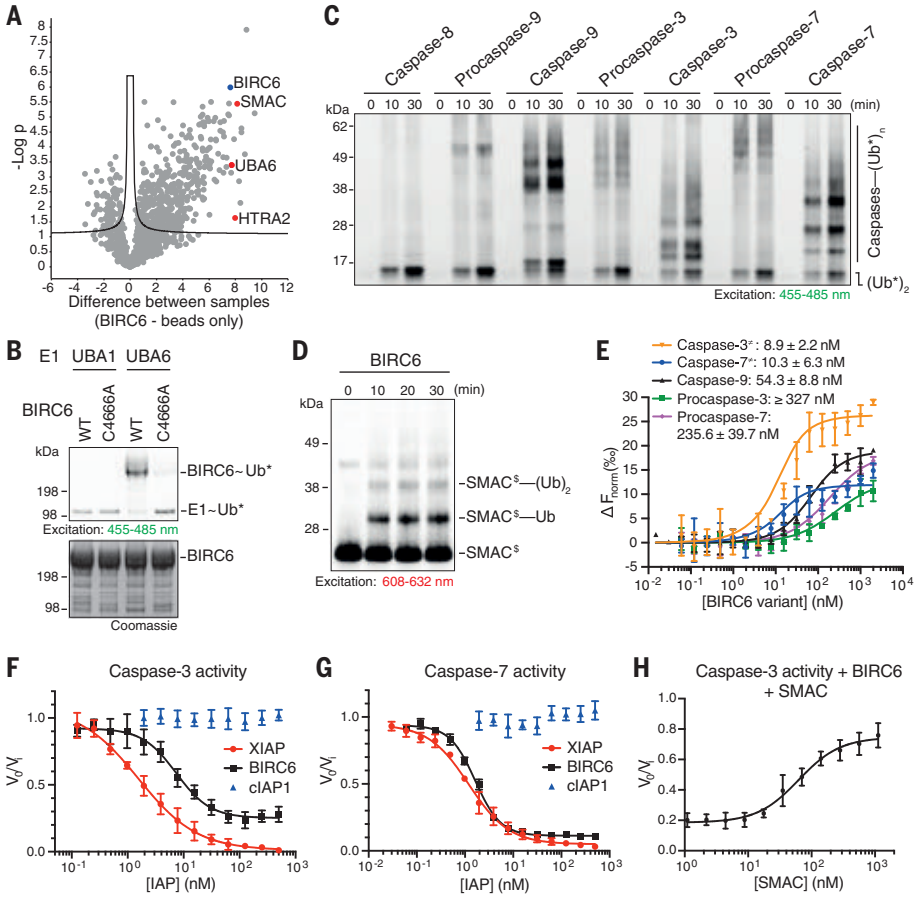


Fig. 1. BIRC6 is a regulator of cell death by direct and indirect inhibition of caspases. (**A**) Identification of BIRC6 interaction partners using affinity-purified tandem mass spectrometry (MS/MS). A volcano plot depicts significant enrichment of proteins bound to BIRC6 compared with a beads-only control. (**B**) BIRC6 receives activated ubiquitin from UBA6 and not UBA1 in an in vitro transthiolation reaction analyzed on a nonreducing SDS–polyacrylamide gel electrophoresis (SDS-PAGE) gel. C4666A, catalytically inactive mutant; ~, thioester bond. (**C**) In vitro ubiquitination by BIRC6 of initiator and executioner caspases as nonprocessed (Pro-) or activated forms using BODIPY (BDP)–labeled ubiquitin (Ub*). (**D**) BIRC6 ubiquitinates a key regulator of intrinsic apoptosis, SMAC, shown through an in vitro ubiquitination assay as in (C) but using Cy5-labeled SMAC ($). –, covalent bond. (**E**) BIRC6 binds activated caspases with high affinity, measured by microscale thermophoresis (MST). ≠, use of activated caspase devoid of catalytic activity. ΔF_{norm}, change in normalized fluorescence; ‰, per mil. (**F** and **G**) BIRC6 directly inhibits activity of caspase-3 (F) and caspase-7 (G) in a fluorogenic substrate cleavage assay. Graphs show rate of substrate cleavage by caspase-3 or -7 in the presence of IAPs, normalized to activity in the absence of IAPs. (**H**) SMAC alleviates BIRC6-mediated caspase inhibition in a fluorogenic substrate cleavage assay recording caspase-3 activity in the presence of 25-fold molar excess BIRC6 to caspase-3 with varying SMAC concentrations normalized to caspase activity alone. Gels and graphs are representative of three independent repeats. Error bars represent standard deviations (SDs).

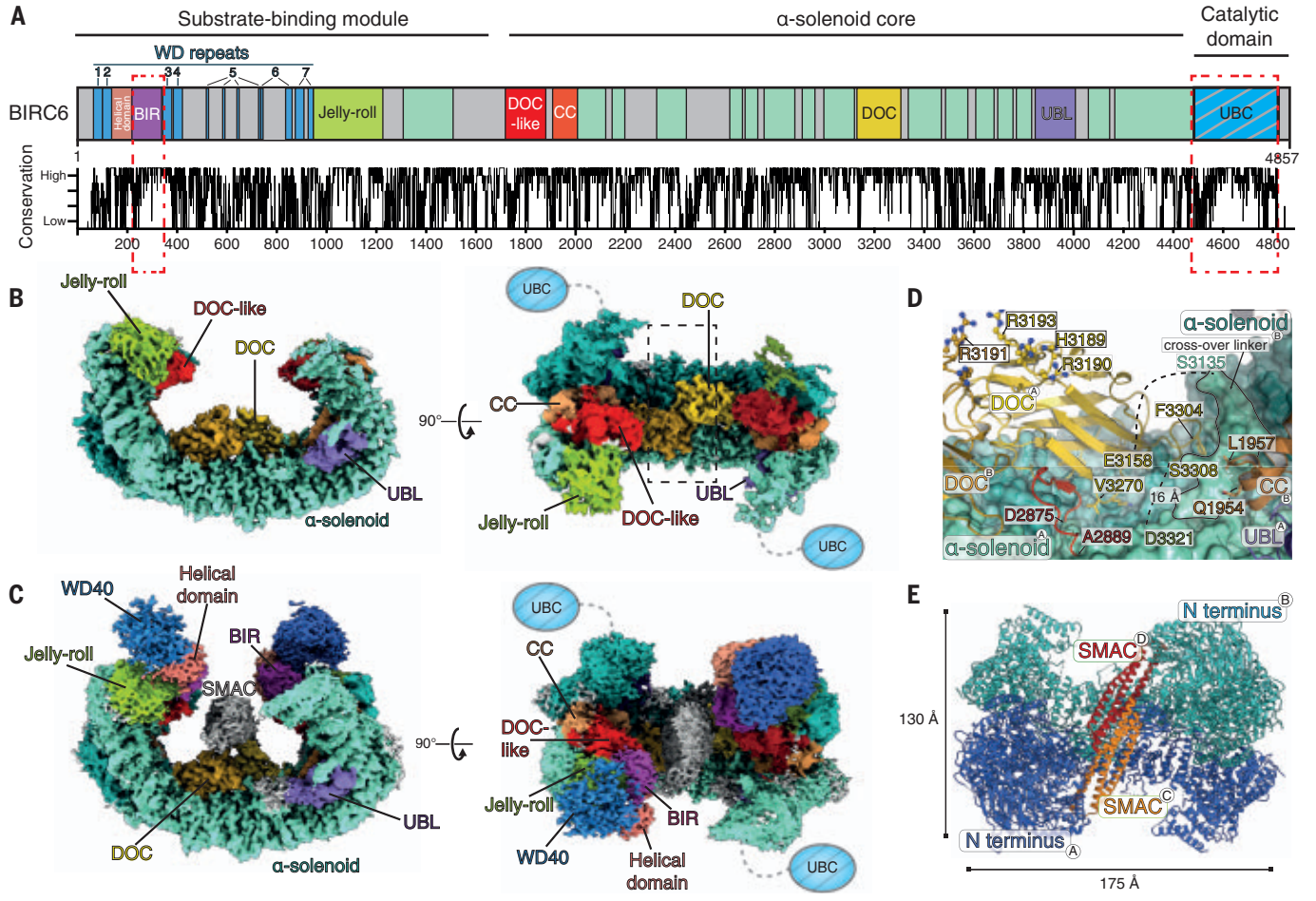


Fig. 2. The antiparallel dimeric architecture and intricate domain arrangement of BIRC6 is important for SMAC engagement. (**A**) Domain annotation of BIRC6 revealed through our study. Corresponding residue conservation of BIRC6 among 126 orthologs is shown below. Previously known BIR and UBC domains are highlighted in red dashed boxes. Areas in gray are not visible in our cryo-EM structures. CC, coiled-coil. (**B**) Composite cryo-EM density representing BIRC6 contoured at 3σ. Domains are colored according to (A). Black dashed box highlights enlarged region in (D). (**C**) BIRC6 in complex with SMAC shown by cryo-EM density contoured at 2σ. Molecular dynamics flexible fitting was used to fit AlphaFold2 (*44*) models of SMAC, the WD40, BIR, and helical domains into the low-threshold regions of the map. Domains are colored according to (A). (**D**) Magnification of the central DOC domain residues Ile^{3160} to Ser^{3308} (yellow) showing the distinctive protruding loop and interchain crossover from a side view with the UBL (purple) and coiled-coil (bronze) domains to the right. Chains to which the domains belong are indicated by the circled letters A and B. Red denotes residues Asp^{2875} to Ala^{2889} rising from the α-solenoid core to stabilize the DOC domain positioned over the adjacent chain. Dashed lines indicate flexible loops between DOC and α-solenoid domains. The loop between Ser^{3308} and Asp^{3321} bridges 16 Å. (**E**) Cartoon representation of BIRC6 dimer looking into the central cavity with chains A and B colored blue and cyan, respectively, bound to dimeric SMAC with chains C and D colored orange and red, respectively. 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.

in the central cavity (Fig. 2, A and B). Toward the C-terminal arm, we identify a UBL domain containing a long unstructured loop insertion (fig. S6, C and D). Of central importance to the structural arrangement of these domains is an antiparallel coiled-coil running the length of the arm, nestled between the antiparallel solenoid core and making contacts with the DOC-like and UBL domains, the latter from the adjacent chain (fig. S6, E to G). Additionally, the elbow of the coiled-coil (Gln^{1954} and Leu^{1957}) forms contacts to the helix, leading to the DOC domain. Together, the coiled-coil thus acts as a structural hinge stabilizing the intricate interwoven domain arrangement (fig. S6, F and G).

Our SMAC-bound BIRC6 structure revealed domains in the N-terminal substrate-binding module, including a WD40 domain formed over 1000 residues interrupted by a short helical domain distinct to BIRC6 and followed by a BIR domain (fig. S7, A to D). This interrupted WD40-BIR domain assembly forms a tightly packed structural element and further expands its repertoire of potential substrate interaction regions (Fig. 2A).

We extended our analysis to other BIRC6 orthologs, which revealed that all these structural features are conserved (figs. S8 and S9). This supports the conclusions that the overall architecture and domain composition is essential for BIRC6 function.

Comparing our BIRC6 alone and SMAC-bound structures, small, subtle differences [root mean square deviation (RMSD) of 2.8 Å] are observed across the central cavity and surrounding arms (residues 1307 to 4502) with very little conformational difference in the solenoid core and DOC domain (residues 2100 to 3600; RMSD value of 1.6 Å) (fig. S7, G and H). Therefore, SMAC binding does not induce pronounced conformational changes in the BIRC6 core.

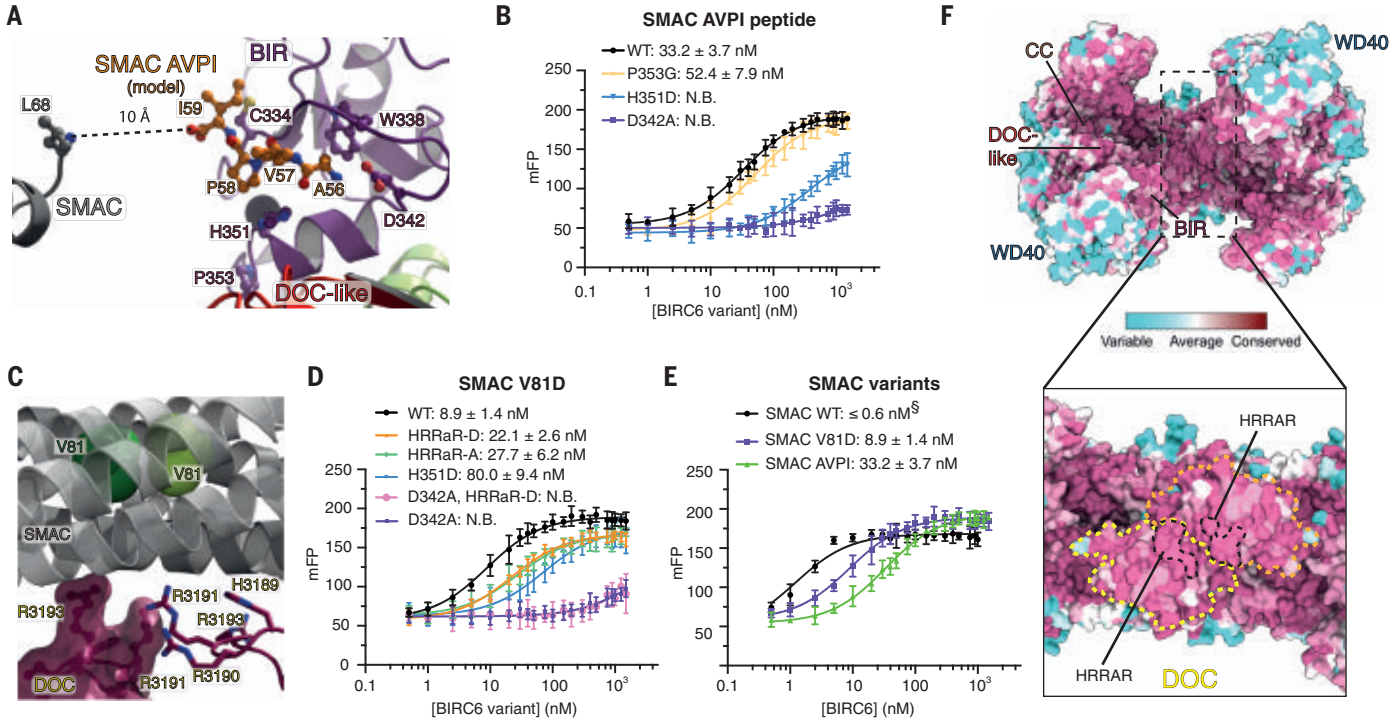


Fig. 3. Dimeric SMAC binds to BIRC6 through a multisite mechanism. (**A**) Position of SMAC N-terminal peptide (orange) binding to BIRC6 BIR domain (purple) modeled by molecular dynamics modeling. The peptide C terminus is 10 Å from the N terminus of the SMAC helical bundle (Leu^{68}), providing sufficient distance for seven amino acids to span. (**B**) Affinities of BIRC6 WT and variants to Cy3-labeled SMAC N-terminal peptide, determined using FP. (**C**) The DOC domain of BIRC6 provides an additional interface for SMAC binding involving a highly conserved, positively charged sequence at the tip of the DOC domain (residues 3189 to 3193; HRRAR). (**D**) Affinity measurements between Alexa 555 monomeric SMAC (V81D) and BIRC6 variants determined by FP. N.B., no binding. (**E**) Determination of BIRC6 affinity to labeled SMAC (WT), SMAC monomer (V81D), and SMAC peptide by FP. §, BIRC6-SMAC (WT) affinity being tighter than the minimum concentration of labeled SMAC detectable. (**F**) Conservation analysis of 126 BIRC6 orthologs with a magnification of the highly conserved central DOC domain outlined in yellow and the HRRAR tip outlined in black. Highly conserved residues are shown in maroon and variable residues in cyan, as defined by ConSurf (*45*). Graphs are representative of three independent repeats. Error bars represent SDs.

Instead, binding of SMAC restricts the movement of the highly flexible N-terminal substrate-binding module, which allows visualization of this additional region.

SMAC binding to BIRC6

We examined the interactions between BIRC6 and SMAC in more detail. Our BIRC6-SMAC structure shows dimeric SMAC positioned above the DOC domain (Fig. 2C and fig. S7, A and E). The start of the SMAC helical bundle (Leu^{68}) points toward the BIR domain (Fig. 3A), which indicates a potential interaction with the N terminus of mature SMAC. Through molecular docking simulations, the SMAC N-terminal AVPI sequence binds to a well-defined pocket in the BIRC6 BIR domain with an invariant residue Asp^{342} hydrogen bonding to SMAC N-terminal Ala^{56} (Fig. 3A), analogous to the reported XIAP BIR2 and BIR3 SMAC interactions (*30*, *31*) (fig. S10A). In support of this model, we recorded a strong, nanomolar interaction between BIRC6 and the SMAC N-terminal peptide using fluorescence polarization (FP) and mutation of Asp^{342} or His^{351} in full-length BIRC6 abrogated binding to the SMAC peptide (Fig. 3B and fig. S10B); mutating Pro^{353} positioned on the external face of the peptide-binding pocket did not affect peptide-binding affinity (Fig. 3B). In conclusion, the N- terminal residues of SMAC bind to the BIR domain through the conserved Asp^{342}-containing pocket.

BIRC6 binding to monomeric SMAC (V81D) is approximately four times as strong (K_d of 8.9 ± 1.4 nM) as binding to the isolated SMAC peptide (K_d of 33.2 ± 3.7 nM) (Fig. 3D) and is driven through the Asp^{342}-containing pocket (loss of Asp^{342} abrogated binding) (Fig. 3D). Consistent with our complex structure, we detect a binding contribution from the DOC domain protruding loop because mutating the positively charged residues in the DOC domain tip to all Ala (HRRaR-A) or all Asp (HRRaR-D) reduced the binding affinity comparable to that measured between BIRC6 and the N-terminal peptide of SMAC (Fig. 3, B to D, and fig. S7F).

Dimeric, wild-type (WT) SMAC binds BIRC6 with a subnanomolar affinity estimated through a competition assay with the SMAC peptide (Fig. 3E and fig. S10C), in agreement with the SMAC-XIAP BIR2-BIR3 affinity reported (*32*). Mutation of D342A in combination with HRRaR-D reduced, but did not abolish, binding to dimeric SMAC (fig. S10D), which suggests that SMAC dimerization strengthens additional weaker multivalent interactions with BIRC6 not observed in our structure.

Together, our biophysical analyses support that dimeric SMAC binding to dimeric BIRC6 is multisite, driven through the BIR domain with additional contributions from the DOC domain and other interfaces. Conservation analyses among a large range of BIRC6 orthologs reveal a highly invariant inner cavity incorporating the DOC domain with a conserved protruding loop (Fig. 3F and fig. S9), which indicates essentiality for substrate recognition in the mouth of BIRC6.

SMAC antagonizes BIRC6-caspase binding

We then investigated what role the BIRC6-SMAC binding mechanism plays in the inhibition of caspases. First, we tested whether BIRC6-caspase binding involves the same key domains and residues as SMAC binding. BIR domain D342A mutation substantially reduced or even abrogated BIRC6 binding to activated caspase-3, -7, and -9 (Fig. 4, A and B, and fig.

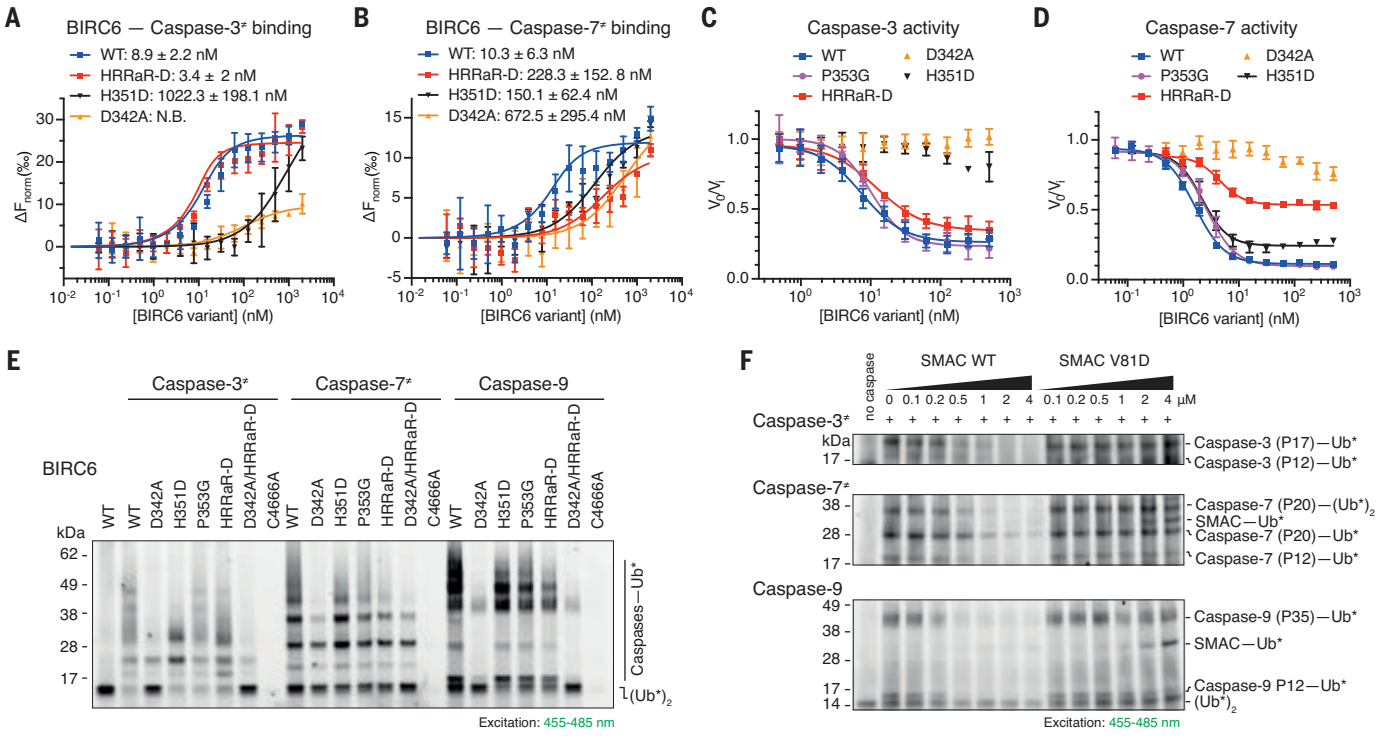


Fig. 4. Caspases and SMAC bind to BIRC6 through overlapping sites. (**A** and **B**) Affinities between BIRC6 variants and RED-tris-NTA–labeled activated caspase-3 (A) and caspase-7 (B) measured using MST. ≠, activated caspase devoid of catalytic activity. (**C** and **D**) Fluorogenic substrate cleavage assays testing ability of BIRC6 variants to inhibit caspase-3 (C) or caspase-7 (D). Plots indicate amount of substrate cleaved by active caspase in the presence of BIRC6 variants normalized against caspase activity alone. (**E**) In vitro ubiquitination assays using BDP-labeled ubiquitin (Ub*) comparing the effect of BIR domain and DOC domain mutations on BIRC6 ubiquitination of activated caspases. (**F**) Dimeric SMAC impairs BIRC6 ubiquitination of caspase-3, -7, and -9. Increasing concentrations of SMAC dimer (WT) or monomer (V81D) were incubated with BIRC6 before the addition of caspase-3(≠), caspase-7(≠), or caspase-9 in an in vitro ubiquitination reaction using Ub*. Gels and graphs are representative of three independent repeats. Error bars represent SDs.

S10E). Crucially, BIRC6 D342A no longer inhibited caspase-3 and -7 activity (Fig. 4, C and D) and displayed reduced ubiquitination of caspase-3, -7, and -9 (Fig. 4E).

Notably, similarly to SMAC-BIRC6 binding, caspases interact with BIRC6 through multivalent mechanisms. Whereas caspase-3, -7, and -9 binding is driven through the BIR domain invariant Asp^{342}, we recorded an additional contribution from the DOC domain in the BIRC6–caspase-7 interaction—mutating the DOC domain tip (HRRaR-D) alone markedly reduced BIRC6–caspase-7 binding and inhibition (Fig. 4, B and D) while not distinctly affecting BIRC6–caspase-3 interaction (Fig. 4, A and C). Furthermore, caspase-3 binding was more dependent on residues lining the BIR domain pocket because mutation of His^{351} substantially reduced caspase-3 affinity and abrogated caspase-3 inhibition. By comparison, BIRC6 H351D–caspase-7 affinity was less affected, and inhibition was retained (Fig. 4, A to D) owing to a greater binding contribution from the DOC domain. Together, this indicates differential mechanisms of binding between BIRC6 and caspase-3 and -7.

Finally, mutating Pro^{353}—corresponding to XIAP Pro^{325}, important for XIAP–caspase-9 binding (*33*) but positioned away from the BIRC6 SMAC peptide-binding pocket (Fig. 3A and fig. S10A)—did not affect BIRC6–caspase-9 binding or ubiquitination (Fig. 4E and fig. S10E) or inhibition of caspase-3 and -7 (Fig. 4, C and D). This indicates that BIRC6 inhibition of caspases differs from XIAP-mediated inhibition of caspase-9.

Therefore, not only are the BIR and DOC domains also involved in BIRC6 inhibition of caspase-3, -7, and -9, but, critically, residues pivotal for caspase binding are identical to those bound by SMAC. This suggests that SMAC directly competes with these caspases for BIRC6 binding.

Lastly, we assessed whether the affinity of SMAC binding to BIRC6 contributes to its antagonism. Dimeric SMAC binds BIRC6 with at least 10 times greater affinity than that of monomeric SMAC (V81D) (Fig. 3E). Whereas WT SMAC antagonizes BIRC6 ubiquitination of caspase-3, -7, and -9, monomeric SMAC (V81D) notably showed impaired ability to inhibit BIRC6 caspase ubiquitination, even at higher SMAC concentrations (Fig. 4F and fig. S10, F to H). Therefore, the multisite binding of dimeric SMAC to dimeric BIRC6 that we observe in our cryo-EM structure results in a subnanomolar affinity that is key for the ability of SMAC to block BIRC6 antiapoptotic function.

Implications for BIRC6 function

These findings cement BIRC6 as a bona fide caspase inhibitor. BIRC6 physically restricts activity of caspase-3 and -7, positioning BIRC6 alongside XIAP as the only IAPs with this behavior (*33*–*36*). Furthermore, BIRC6 combines this inhibitory activity with that of cIAP1/2 through ubiquitinating activated caspases (*6*, *11*, *37*); BIRC6 may collude with another E3 ubiquitin ligase, extending the initially deposited monoubiquitin and leading to caspase degradation in a cellular context (fig. S11).

We uncover that SMAC antagonizes BIRC6 through directly outcompeting caspases bound to key residues in conserved domains. SMAC dimerization and the constitutive antiparallel, dimeric nature of BIRC6 are critical for this activity through a resultant subnanomolar affinity complex. Our structures also reveal an additional antagonistic role of the SMAC central helical bundle in IAP inhibition. In addition to it enabling SMAC dimerization—key for simultaneous binding of BIR2 and BIR3 domains within XIAP (*20*, *21*) and cIAP1/2 and

critical for engagement of the resultant two BIRC6 BIR domains revealed through our structure—interactions of the SMAC helical bundle with the BIRC6 DOC domain further increase the interaction affinity and physically obstruct other substrates from binding in the highly conserved central cavity.

IAPs, including BIRC6, are frequently overexpressed in cancers (*38*–*41*), and detailed understanding of XIAP and cIAP1/2 inhibitory activity and regulation has enabled the design of therapeutics exploiting the ability of SMAC to block XIAP and cIAP1/2 function to restore apoptosis (*42*). Before now, the BIRC6 BIR domain was thought to be divergent from those of XIAP and cIAP1/2 (*43*), precluding it from being a target for the design of small-molecule inhibitors. We reveal that the SMAC peptide binds the BIRC6 BIR domain in a near-identical manner to XIAP BIR2 and BIR3 (fig. S10A). Our findings illuminate the mechanisms of BIRC6 antiapoptotic activity and multimodal antagonism by SMAC and provide a biochemical and structural framework for future design of small-molecule inhibitors targeting this IAP.

Finally, the discovery of BIRC6 UBA6-dependent ubiquitination activity opens the way for exploring BIRC6 control of emerging UBA6-regulated pathways (*16*, *23*), potentiated by the array of highly conserved substrate-recruitment domains that we identify. Its ability to function as a stand-alone E2 in addition to a dual E2 and E3 ubiquitin ligase further extends the reach of BIRC6 through cooperating with distinct E3 ubiquitin ligases.

REFERENCES AND NOTES

1. S. Kesavardhana, R. K. S. Malireddi, T.-D. Kanneganti, *Annu. Rev. Immunol.* **38**, 567–595 (2020).
2. M. Hrdinka, M. Yabal, *Genes Immun.* **20**, 641–650 (2019).
3. B. P. Eckelman, G. S. Salvesen, F. L. Scott, *EMBO Rep.* **7**, 988–994 (2006).
4. M. J. M. Bertrand *et al.*, *Mol. Cell* **30**, 689–700 (2008).
5. J. Zhang *et al.*, *Cell Rep.* **27**, 2679–2689.e3 (2019).
6. Y. E. Choi *et al.*, *J. Biol. Chem.* **284**, 12772–12782 (2009).
7. K. Lotz, G. Pyrowolakis, S. Jentsch, *Mol. Cell. Biol.* **24**, 9339–9350 (2004).
8. J. Ren *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 565–570 (2005).
9. K. Yanagitani, S. Juszkiewicz, R. S. Hegde, *Science* **357**, 472–475 (2017).
10. T. Bartke, C. Pohl, G. Pyrowolakis, S. Jentsch, *Mol. Cell* **14**, 801–811 (2004).
11. Y. Hao *et al.*, *Nat. Cell Biol.* **6**, 849–860 (2004).
12. A. Lopergolo *et al.*, *Br. J. Cancer* **100**, 739–746 (2009).
13. C. Domingues, H. D. Ryoo, *Cell Death Differ.* **19**, 470–477 (2012).
14. P. Ebner *et al.*, *Nat. Commun.* **9**, 599 (2018).
15. J. J. Gilchrist *et al.*, *eLife* **11**, e77461 (2022).
16. R. Jia, J. S. Bonifacino, *eLife* **8**, e50034 (2019).
17. C. Pohl, S. Jentsch, *Cell* **132**, 832–845 (2008).
18. S. M. Srinivasula *et al.*, *J. Biol. Chem.* **278**, 31469–31472 (2003).
19. Q.-H. Yang, R. Church-Hajduk, J. Ren, M. L. Newton, C. Du, *Genes Dev.* **17**, 1487–1496 (2003).
20. J. Chai *et al.*, *Nature* **406**, 855–862 (2000).
21. A. M. Verhagen *et al.*, *Cell* **102**, 43–53 (2000).
22. X.-B. Qiu, A. L. Goldberg, *J. Biol. Chem.* **280**, 174–182 (2005).
23. X. Liu *et al.*, *Nat. Commun.* **8**, 14286 (2017).
24. J. Jin, X. Li, S. P. Gygi, J. W. Harper, *Nature* **447**, 1135–1138 (2007).
25. Y.-H. Chiu, Q. Sun, Z. J. Chen, *Mol. Cell* **27**, 1014–1023 (2007).
26. M. Zhou *et al.*, *Genes Dev.* **29**, 2349–2361 (2015).
27. P. C. A. da Fonseca *et al.*, *Nature* **470**, 274–278 (2011).
28. D. Barford, *Curr. Opin. Struct. Biol.* **61**, 86–97 (2020).
29. A. H. Tencer *et al.*, *Sci. Rep.* **12**, 6063 (2022).
30. G. Wu *et al.*, *Nature* **408**, 1008–1012 (2000).
31. C. Lukacs *et al.*, *Acta Cryst.* **D69**, 1717–1725 (2013).
32. Y. Huang, R. L. Rich, D. G. Myszka, H. Wu, *J. Biol. Chem.* **278**, 49517–49522 (2003).
33. E. N. Shiozaki *et al.*, *Mol. Cell* **11**, 519–527 (2003).
34. S. J. Riedl *et al.*, *Cell* **104**, 791–800 (2001).
35. S. M. Srinivasula *et al.*, *Nature* **410**, 112–116 (2001).
36. J. Chai *et al.*, *Cell* **104**, 769–780 (2001).
37. Hk. Huang *et al.*, *J. Biol. Chem.* **275**, 26661–26664 (2000).
38. Z. Chen *et al.*, *Biochem. Biophys. Res. Commun.* **264**, 847–854 (1999).
39. X. Dong *et al.*, *J. Thorac. Oncol.* **8**, 161–170 (2013).
40. T. Hu *et al.*, *PLOS ONE* **10**, e0125281 (2015).
41. Y. Li *et al.*, *Mol. Ther. Nucleic Acids* **26**, 798–812 (2021).
42. E. Morrish, G. Brumatti, J. Silke, *Cells* **9**, 406 (2020).
43. M. P. Schwalm *et al.*, *Front. Cell Dev. Biol.* **10**, 886537 (2022).
44. K. Tunyasuvunakool *et al.*, *Nature* **596**, 590–596 (2021).
45. M. Landau *et al.*, *Nucleic Acids Res.* **33**, W299–W302 (2005).

ACKNOWLEDGMENTS

We acknowledge A. von Kugelgen, L. Carrique, T. Matthews-Palmer, A. Costin, and J. Caesar for help and advice with cryo-EM data collection and processing. Electron microscopy provision was provided through eBIC (proposal BI28731) and the Central Oxford Structural Molecular Imaging Centre (COSMIC) electron microscopy facility. Mass spectrometry analysis was performed in the Discovery Proteomics Facility (Target Discovery Institute) led by R. Fischer and I. Vendrell. We thank F. Barr, M. Higgins, and S. Newstead for critical reading of the manuscript. **Funding:** L.D. and C.R. are funded through a Wellcome Trust studentship in Cellular and Structural Biology. A.P.-F. and B.M.K. are supported by the Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Science (CIFMS), China (grant no. 2018-I2M-2-002), and by Pfizer. P.R.E. is supported by a Medical Research Council (MRC) career development fellowship (MR/R008582/1). **Author contributions:** P.R.E. designed and initiated the study. L.D. performed cryo-EM sample preparation, data collection, and analysis. L.D., C.J.E., C.R., F.D.F., and P.R.E. performed experiments. Molecular dynamics simulations and flexible fitting were performed by C.K.C. A.P.-F. performed peptide in-solution digest and data analysis, supervised by B.M.K. P.R.E. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Cryo-EM maps are deposited in the Electron Microscopy Data Bank under accession numbers EMD-15648, 15650, 15651, 15652, 15653, and 15654. Atomic coordinates were deposited in the Protein Data Bank (PDB) under accession numbers 8ATM and 8ATO for the BIRC6 and BIRC6-SMAC structures, respectively. Full mass spectrometry data is available through the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD039954.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade8840
Materials and Methods
Figs. S1 to S11
Table S1
References (*46*–*71*)
MDAR Reproducibility Checklist
Data S1

Submitted 14 September 2022; accepted 31 January 2023
10.1126/science.ade8840

STRUCTURAL BIOLOGY

Structural basis for regulation of apoptosis and autophagy by the BIRC6/SMAC complex

Julian F. Ehrmann[1,2]*†, Daniel B. Grabarczyk[1]*†, Maria Heinke[1], Luiza Deszcz[1], Robert Kurzbauer[1], Otto Hudecz[1], Alexandra Shulkina[2,3], Rebeca Gogova[1,2], Anton Meinhart[1], Gijs A. Versteeg[3], Tim Clausen[1,4]*

Inhibitor of apoptosis proteins (IAPs) bind to pro-apoptotic proteases, keeping them inactive and preventing cell death. The atypical ubiquitin ligase BIRC6 is the only essential IAP, additionally functioning as a suppressor of autophagy. We performed a structure-function analysis of BIRC6 in complex with caspase-9, HTRA2, SMAC, and LC3B, which are critical apoptosis and autophagy proteins. Cryo–electron microscopy structures showed that BIRC6 forms a megadalton crescent shape that arcs around a spacious cavity containing receptor sites for client proteins. Multivalent binding of SMAC obstructs client binding, impeding ubiquitination of both autophagy and apoptotic substrates. On the basis of these data, we discuss how the BIRC6/SMAC complex can act as a stress-induced hub to regulate apoptosis and autophagy drivers.

Apoptosis is an evolutionarily conserved and essential form of programmed cell death used to remove damaged or surplus cells (*1*). Dysregulation of apoptosis can lead to disease, in particular cancer, atrophy, and neurodegenerative disorders (*2*, *3*). To safeguard against the untimely initiation of apoptosis, a family of functionally related proteins called inhibitor of apoptosis proteins (IAPs) bind and inhibit caspases, cysteine proteases that execute the cell death program (*4*). A common feature of all IAPs is the presence of a BIR domain, which binds to a specific N-terminal signal sequence called an N-degron in its target proteins (*4*). Active

[1]Research Institute of Molecular Pathology, Vienna BioCenter, Vienna, Austria. [2]Vienna BioCenter PhD Program, Doctoral School of the University of Vienna and Medical University of Vienna, Vienna BioCenter, Vienna, Austria. [3]Max Perutz Labs, University of Vienna, Vienna BioCenter, Vienna, Austria. [4]Medical University of Vienna, Vienna, Austria.
***Corresponding author. Email: julian.ehrmann@imp.ac.at (J.F.E.); daniel.grabarczyk@imp.ac.at (D.B.G.); tim.clausen@imp.ac.at (T.C.)**
†These authors contributed equally to this work.

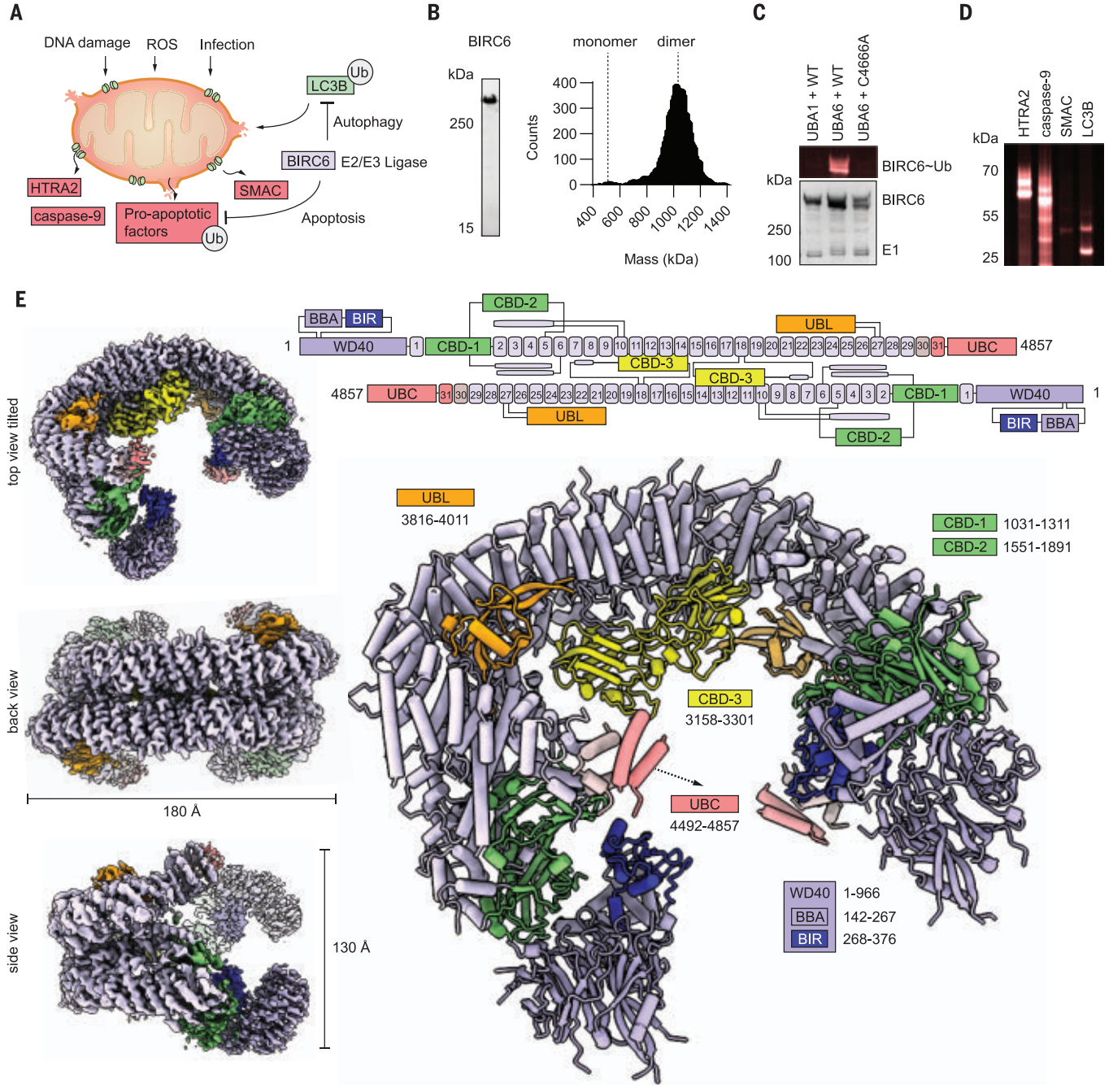


Fig. 1. Cryo-EM structure of human BIRC6 in its active dimeric form. (**A**) BIRC6 inhibits apoptosis and autophagy as an E2/E3 hybrid, and ubiquitinated targets are marked for subsequent proteasomal degradation. (**B**) Recombinant BIRC6 purity was assessed by SDS–polyacrylamide gel electrophoresis, and oligomeric state was determined by mass photometry measurements at 20 nM. (**C**) Auto-ubiquitination activity of BIRC6 assessed with the E1 enzymes UBA1 or UBA6 with wild-type or catalytically inactive BIRC6 C4666A. DyLight488-labeled ubiquitin is spiked into the reaction for in-gel visualization (top). (**D**) Ubiquitination assay of reported interaction partners HTRA2, caspase-9, SMAC, and LC3B. Gels and ubiquitination assays are representative of at least three experimental replicates. (**E**) Cryo-EM map of the BIRC6 homodimer, present in C2 symmetry, in three orientations with dimensions indicated (left) and the model in cartoon representation (right). Domains are annotated with respective sequence boundaries and a schematic of the dimeric assembly is shown (top).

caspases cleave themselves to expose internal N-degrons as the basis for IAP recognition. During apoptosis, damaged mitochondria release the effector protein SMAC, which also presents an N-degron and competes for BIR domain binding, thereby liberating caspases from IAPs (*5*, *6*). Acting within the ubiquitin cascade of E1-activating, E2-conjugating, and E3-ligating enzymes, most IAPs contain a RING domain that has E3 ubiquitin ligase activity (*7*). This activity covalently modifies the target substrate with the small protein ubiquitin, directing it for degradation by the proteasome. Thus, binding of the IAP to its target both inhibits pro-apoptotic proteases and promotes their degradation.

The human IAP BIRC6 (also known as BRUCE or APOLLON) is the only essential IAP family member (*8*, *9*). Despite its giant size of 530 kDa, it has only two characterized domains, a BIR domain for N-degron binding and a UBC

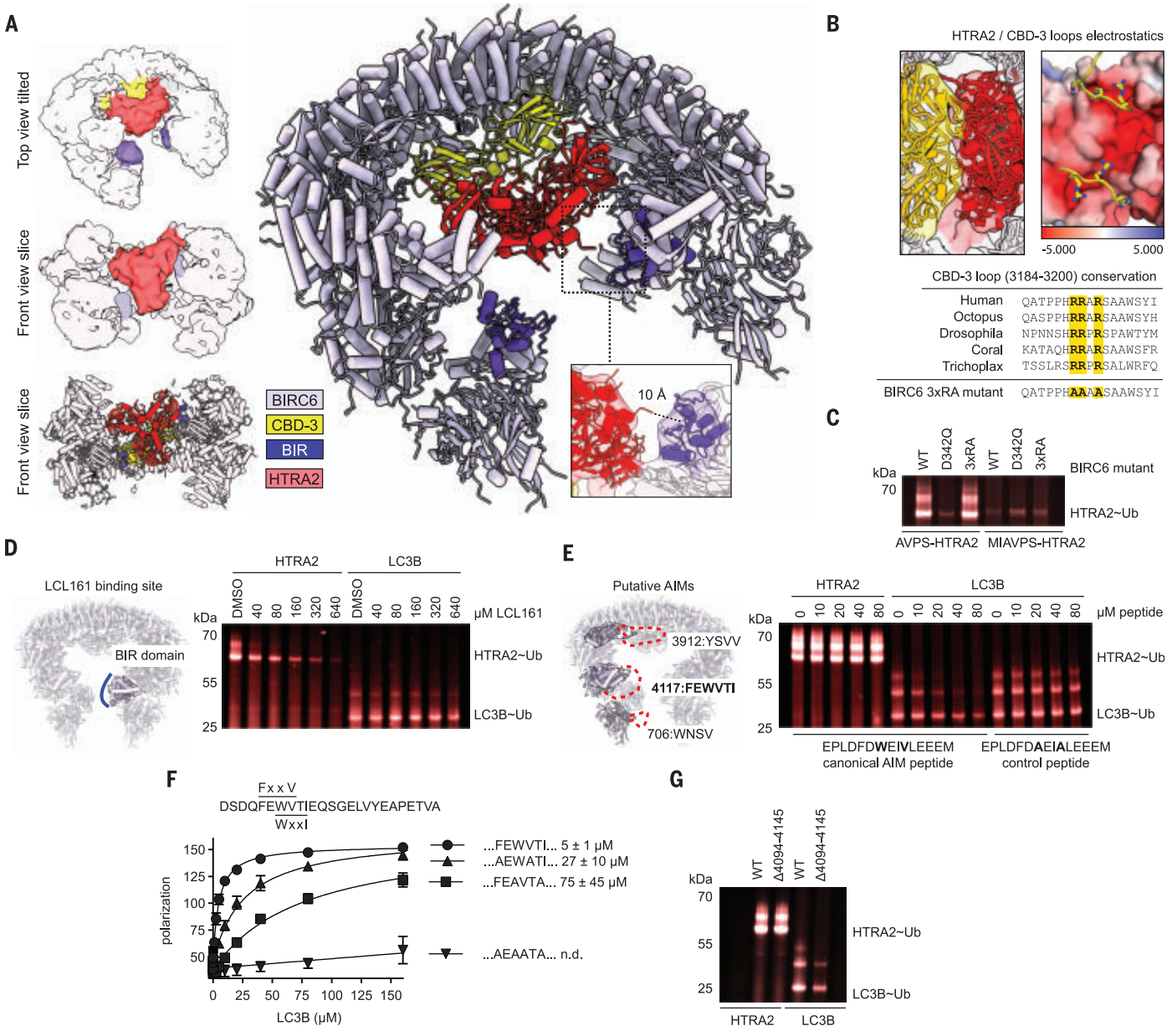


Fig. 2. Modes of HTRA2 and LC3B substrate recruitment. (**A**) Cryo-EM structure of BIRC6 bound to HTRA2 low-pass filtered to 10 Å (left) is shown together with a cartoon representation of the E3:substrate complex highlighting the distance of the N-degron termini to the BIR domain (center). (**B**) Cartoon and electrostatic potential (± 5 $k_BTe_c^{-1}$ at 298.15 K) representation of the CBD-3 arginine loops packing against the negatively charged protease. Representative alignment highlighting the three conserved arginine residues used to construct the 3xRA mutant (Arg3190Ala/Arg3191Ala/Arg3193Ala) is shown below. Octopus, *Octopus sinsensis*; Drosophila, *Drosophilamelanogaster*; Coral, *Orbicella faveolate*; Trichoplax, *Trichoplax* sp. *H2*. (**C**) Ubiquitination assays of wild-type and BIR domain mutated BIRC6 (D342Q) against HTRA2 with a native N-degron (AVPS) or a variant (MIAVPS) unable to bind the BIRC6 BIR domain. (**D**) HTRA2 and LC3B ubiquitination when titrating in an increasing concentration of the IAP inhibitor LCL161, a BIR domain ligand. (**E**) HTRA2 and LC3B ubiquitination when titrating in an increasing concentration of an AIM peptide or a control peptide with the ATG8-interacting motif disrupted. (**F**) Fluorescence polarization assays of LC3B binding to a peptide of BIRC6 AIM-4117–derived motif (sequence shown) and its mutants. (**G**) Ubiquitination assays of a BIRC6 construct with a perturbation in the ATG8 binding region (residues Δ4094 to 4145) compared with wild type. All assays were performed as technical triplicates.

domain at its C terminus, which is characteristic of E2 ubiquitin–conjugating enzymes. The UBC domain of BIRC6 is homologous to those of UBE2O and UBE2Z, which function as hybrid E2/E3 enzymes mediating ubiquitin transfer without requiring a cognate E3 ligase (*10*–*12*). Both the BIR domain and UBC domain of BIRC6 are important for its role in down-regulating N-degron–containing pro-apoptotic proteases such as caspases and HTRA2 (*13*–*16*). However, this regulation works in both directions, and these proteases also exploit the BIR domain to degrade BIRC6 (*13*, *14*). A distinguishing feature of BIRC6 compared with other IAPs are functions extending beyond apoptosis regulation (*17*–*20*), exemplified by its ability to inhibit autophagy (*17*, *21*) (Fig. 1A). During autophagy, ATG8 family proteins such as LC3B recruit cargo-carrying receptor proteins and form autophagosomal membranes when lipidated (*22*). By ubiquitinating LC3B

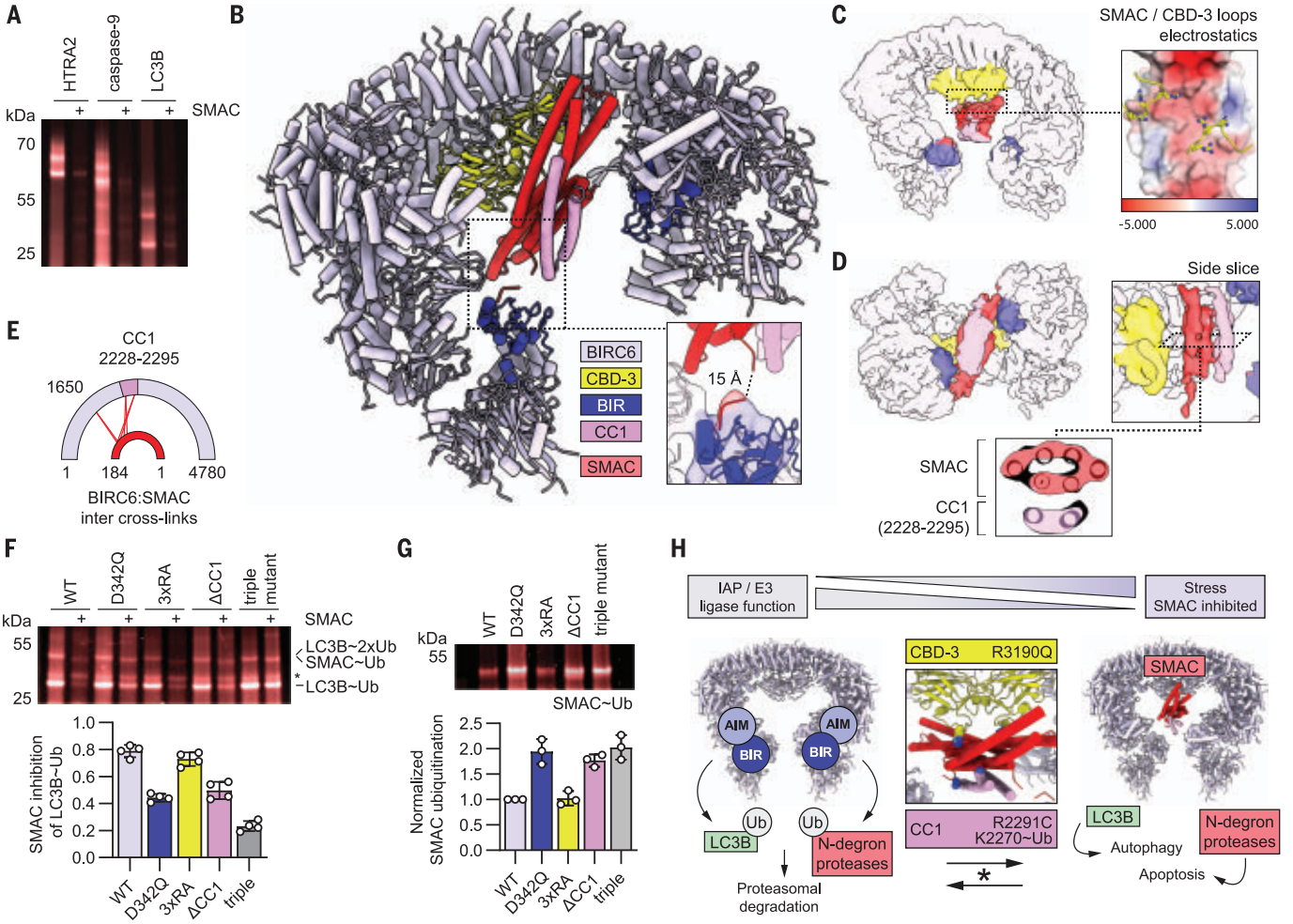


Fig. 3. BIRC6 inhibition by SMAC is dependent on a specific tripartite binding mode. (**A**) Assays showing ubiquitination of HTRA2, caspase-9, and LC3B in the presence and absence of SMAC. Experiment was performed three times as a technical replicate. (**B**) Cryo-EM structure of BIRC6 bound to SMAC. The main panel shows the model in cartoon representation with domain coloring indicated below. Experimental degron density is highlighted in the bottom inset. (**C**) Interaction between SMAC and CBD-3. The inset shows the negatively charged surface of SMAC (± 5 $k_BTe_c^{-1}$ at 298.15 K) binding the CBD-3 loops. (**D**) Experimental map and model shown in multiple orientations to illustrate the densities of SMAC (red) and the BIRC6 coiled-coil motif CC1 (purple, residues 2228 to 2295). (**E**) Representation of the multiple intermolecular cross-links between SMAC and the CC1 insert detected by cross-link-coupled mass spectrometry. (**F**) Using LC3B ubiquitination as a functional readout, BIRC6 mutants were tested for sensitivity to SMAC inhibition. The triple mutant combines D342Q, 3xRA, and the CC1 deletion in a single construct. The fraction of the LC3B-Ub band lost upon SMAC addition is quantified for four replicates for each variant. Mean and SD are shown. (**G**) Ubiquitination of SMAC was tested for BIRC6 variants with disrupted interaction sites. Mean and SD of three replicates are plotted with SMAC ubiquitination normalized to the wild-type reaction. (**H**) Schematic of how BIRC6 uses its BIR domain and AIM to select diverse substrates for ubiquitination and how SMAC shuts down this activity by binding the central cavity. The resolved interaction with the regulator SMAC contextualizes prominent substitutions in cancer (R2291C), ubiquitination sites (K2270), and a murine developmental disorder mutation (R3190Q) in BIRC6.

and causing its proteasomal degradation (*21*), BIRC6 limits the availability of this important autophagic building block.

BIRC6 dimerizes around a central ubiquitination zone

To determine the architecture of BIRC6 and its molecular mechanism, we first expressed the human protein in insect cells. Mass photometry indicated that recombinant BIRC6 exists as a stable dimer with a molecular weight of ~1.1 MDa (Fig. 1B). To test the activity of the dimer, we performed in vitro ubiquitination assays and saw activity in the presence of the obligate E1 ubiquitin–activating enzyme UBA6 (*13*, *21*, *23*), validating the E2/E3 hybrid nature of BIRC6 (Fig. 1C). We next determined the activity against reported substrates with exposed N-degrons. BIRC6 strongly ubiquitinated HTRA2 (residues 134 to 458, S306A) and caspase-9, but had only weak activity against SMAC (residues 56 to 239). We also observed ubiquitination of LC3B, which lacks an N-degron (Fig. 1D). We then performed cryo–electron microscopy (cryo-EM) analysis of BIRC6 alone and in the presence of substrates. These data enabled us to generate a map of full-length BIRC6 with an overall resolution of 3.3 Å and build the structure of the active E2/E3 ligase (figs. S1 to S5 and tables S1 and 2). The structure revealed an intricate antiparallel dimer with four front-facing termini connected by a twinned helical scaffold consisting of 31 armadillo repeats (Fig. 1E and fig. S6A). The resulting C-shaped structure has overall dimensions of 180 × 150 × 130 Å and encloses a central cavity. The N-terminal module is a seven-bladed beta propeller with two inserted adjacent domains (fig. S6, B and C): a BIRC6 BIR-associated

domain (BBA, residues 142 to 267) and the N-degron–binding BIR domain (residues 268 to 376). The beta-propeller is flexibly linked to the start of the helical scaffold, which is initiated by two carbohydrate-binding-like domains, CBD-1 (residues 1031 to 1311) and CBD-2 (residues 1551 to 1891) (fig. S6D). Structurally related domains use a common surface patch to mediate interactions with carbohydrates or proteins, for example, contributing to substrate recruitment in another ubiquitin ligase, the anaphase-promoting complex (*24*). For CBD-1 and CBD-2, these surfaces point to the interior of BIRC6 and may serve as accessory binding sites. A third CBD domain (CBD-3, residues 3158 to 3301) is located at the center of BIRC6, where it homodimerizes through the described interface. A ubiquitin-like domain (UBL, residues 3816 to 4011) accompanies a sharp bend in the armadillo repeat backbone toward the front side, where the C-terminal UBC domain (residues 4492 to 4857) is loosely tethered. Although we struggled to resolve the UBC domain in our EM reconstruction, in two-dimensional (2D) classes and 3D variability analysis of a BIRC6 dataset in the absence of substrates, we observed the appearance of globular density matching the dimensions of the UBC module extending from the C-terminal helical scaffold into the central cavity (fig. S6E). The helical backbones of the two BIRC6 protomers interact intimately with each other, as reflected by the large interface of 9100 Å^2 and the marked stability of the dimer (Fig. 1B). In addition to extensive van der Waals interactions between the stacked armadillo repeats, the dimer is stabilized by the interlocking of CBD-3 in the center of the protein. In addition to forming a prominent protrusion in the central cavity, CBD-3 and an adjacent helix are swapped across the dimer, binding to the armadillo repeat of the other protomer (fig. S6F). Overall, the dimer architecture is critical to orient the catalytic UBC and the substrate-binding BIR domain toward the central cavity. This cavity has a diameter of ~50 Å, providing a spacious ubiquitination zone in reach of the mobile UBC heads.

BIRC6 uses distinct targeting mechanisms for apoptosis and autophagy substrates

To visualize the substrate-targeting mechanism, we reconstituted complexes with the client proteins caspase-9, HTRA2, and LC3B. Cryo-EM data of complexes with caspase-9 and LC3B were inconclusive with regard to substrate densities. We observed heterogeneous densities in the central cavity that were similar to those detected in a BIRC6 dataset in the absence of substrates (fig. S7A) and presumably originated from the flexible UBC domain or other low-occupancy disordered regions. In strong contrast, the cryo-EM map of the complex of BIRC6 with HTRA2 revealed clearly defined features of the bound substrate (Fig. 2A and fig. S7, B to D). The local resolution of HTRA2 was ~8 Å and, because of its distinctive pyramidal shape, the substrate could be unambiguously fit into the density (Fig. 2A and fig. S7, C and D). HTRA2 is a homotrimeric protein composed of a trypsin-like protease, a regulatory PDZ domain, and one N-degron per protomer (*25*). Measuring ~66 Å per side, the trimeric protease core almost fully occupies the central cavity of BIRC6. The base of the trimer, where the proteolytic sites are located, packs against the two CBD-3 domains (Fig. 2B and fig. S7E). Both neighboring CBD-3s interact with HTRA2 using the same feature, an extended loop with three highly conserved arginine residues (Fig. 2B). These arginines recognize the charged surface area surrounding the active sites of HTRA2. In addition to undergoing electrostatic interactions with CBD-3, the protease is well positioned for its N-degrons to interact with the BIRC6 BIR domains, further anchoring the substrate to BIRC6. Although we did not detect density for the N-degron, the six N-terminal residues would be able to bridge the 10-Å distance to reach the proximal BIR domain, supporting a stable yet flexible binding mode.

To understand the contribution of the distinct interaction sites toward substrate binding and ubiquitination, we generated BIRC6 variants with either the N-degron–binding site mutated (D342Q) or the CBD-3 arginine loop modified to negate the positive charge (3xRA). We observed that HTRA2 recognition was mainly dependent on the N-degron interaction, because D342Q abrogated ubiquitination (Fig. 2C). Accordingly, adding the chemical N-degron mimetic LCL161 into the reaction as a competitor or mutating the native N-degron of HTRA2 impaired ubiquitination (Fig. 2D). Although it is not critical for ubiquitination, the interaction with CBD-3 could modulate HTRA2 function. In the previously described HTRA2 resting state, the PDZ domains are closely packed against the protease body, blocking access to the proteolytic sites (*25*). Moreover, nuclear magnetic resonance data indicated that binding of activator peptides leads to PDZ displacement and subsequent protease activation (*26*). In our cryo-EM structure, the BIRC6 CBD-3 domains have displaced the PDZ domains, and density for one of these flexible domains can be detected outside of the trimer core (fig. S7F). Thus, binding to BIRC6 could be directly coupled with protease activation, providing a molecular mechanism of how HTRA2 selectively degrades IAPs upon mitochondrial stress (*14*).

Considering its diverse cellular roles, we were curious how BIRC6 recognizes substrates outside of the apoptosis pathway, especially those lacking an N-degron. The best-characterized nonapoptotic target is the autophagy protein LC3B (*21*). Although it has been shown that BIRC6 negatively regulates autophagy by ubiquitinating and thereby depleting LC3B, the targeting mechanism is unknown. One potential docking site is the BIR domain, which mediates diverse protein-protein interactions, as exemplified by the phospho-dependent binding of histone H3 to survivin/BIRC5 (*27*). To test whether LC3B is also recognized by the BIR domain, we performed competition assays with the N-degron mimetic LCL161. In contrast to HTRA2, we observed no effect on LC3B ubiquitination (Fig. 2D). Because LC3B typically binds to proteins through the hydrophobic residues of ATG8-interacting motifs (AIMs) (*22*), which have the consensus sequence W/F/YxxL/V/I, we performed a bioinformatic motif search and, after filtering for only those in disordered regions, identified several putative AIMs in BIRC6 (*28*) (Fig. 2E). We therefore investigated whether targeting of LC3B may be mediated by such a motif and performed competition experiments with a canonical AIM peptide. Whereas the AIM peptide had no effect on targeting HTRA2, ubiquitination of LC3B was strongly reduced, suggesting that one or more of the putative BIRC6 AIMs recognizes LC3B (Fig. 2E).

To identify which of the putative AIMs interacts with LC3B, we performed fluorescence polarization experiments with BIRC6-derived peptides. Only the highly conserved AIM motif FEWVTI (residues 4117 to 4122, located in a flexible loop attached to the armadillo repeat scaffold; table S2) showed an interaction with LC3B and this could be fit to a K_D of 5 ± 1 μM (Fig. 2F and fig. S8, A and B). Unusually, this AIM consists of two overlapping motifs, FxxVxx and xxWxxI. We observed that both motifs contributed to the interaction, with alanine substitution of all four hydrophobic residues required to fully abrogate the interaction (Fig. 2F). To reveal whether this AIM is responsible for LC3B targeting, we mutated it within full-length BIRC6 and performed ubiquitination assays (Fig. 2G). In agreement with the fluorescence polarization data, deleting the entirety of AIM-4117 (residues Δ4094 to 4145) strongly disrupted LC3B ubiquitination without affecting HTRA2 ubiquitination. Overall, we conclude that BIRC6 has distinct targeting sites for apoptotic and autophagy proteins.

SMAC inhibits BIRC6 by multivalent binding to the ubiquitination cavity

Although SMAC forms a complex with BIRC6 (fig. S9A), the apoptotic protein was only weakly ubiquitinated (Fig. 1D). We reasoned that SMAC could compete, through its N-degron, with HTRA2 and caspase-9 and inhibit their ubiquitination. Indeed, we observed a complete loss of HTRA2 and caspase-9 ubiquitination even when these substrates were in eightfold excess of SMAC (Fig. 3A). Unexpectedly, SMAC

also completely inhibited the activity of BIRC6 against LC3B despite the autophagy protein not binding to the BIR domain.

To better understand the inhibitory mechanism of SMAC, we determined the structure of the BIRC6:SMAC complex by cryo-EM. SMAC was resolved at a local resolution of ~8 Å, enabling placement of dimeric SMAC into the density (Fig. 3B and fig. S9B). Our structure shows that the six helices of the SMAC dimer penetrate the entire cavity of BIRC6, occluding the CBD-3 and BIR domains. Like HTRA2, SMAC is held in the center of the cavity by nonspecific interactions of a negatively charged surface with the arginine loops of the two CBD-3 domains (Fig. 3B). This particular surface on SMAC has previously been reported to interact with another IAP, XIAP/BIRC4 (*29*), pointing to a shared binding mechanism to modulate IAP function (fig. S9C). Both N termini of the SMAC dimer protrude toward the BIR domains, and in this case low-resolution density for the bound N-degron can be observed (Fig. 3B and fig. S9D). An additional density corresponding to two helices was observed packing against the SMAC dimer (Fig. 3B). The extra helices could not be reconciled with the reported tetrameric form of SMAC (*30*). We asked whether flexible segments of BIRC6 may contribute to this density, and used cross-linking coupled mass spectrometry to identify the putative additional SMAC binding site. We detected four specific cross-links between SMAC and a partially disordered insert in the helical backbone of BIRC6 that we named coiled-coil 1 (CC1, residues 2228 to 2295; Fig. 3C). This insert was predicted by Alphafold2 to form two antiparallel helices that fit well to our orphan EM density and has the right linker length to reach the modeled position (fig. S9, E and F). In conclusion, SMAC coordination relies on three main interactions: N-degron binding at the BIR domain, electrostatic interactions with CBD-3, and formation of a helical bundle with CC1.

LC3B is inhibited by SMAC despite not sharing any of the identified SMAC:BIRC6 interaction sites. To understand the mechanism of this inhibition, we mutated each interaction site and monitored the effect of SMAC on BIRC6-mediated LC3B ubiquitination (Fig. 3D). Disruption of the BIR domain with the D342Q mutation did not fully restore LC3B ubiquitination, suggesting that the other two sites contribute to SMAC binding. The 3xRA mutation had little effect on SMAC inhibition, but deleting CC1 caused a moderate loss of inhibition, confirming the importance of this interface. A BIRC6 variant mutated in all three sites showed the strongest loss of SMAC-mediated inhibition (Fig. 3D). We conclude that multivalent interactions arrest SMAC in the central cavity, obstructing the ubiquitination zone and inhibiting BIRC6 activity.

Because SMAC itself is poorly ubiquitinated, we tested whether its tripartite binding mode may be causative by ubiquitinating SMAC in the absence of other substrates using the same BIRC6 interface mutants (Fig. 3E). The D342Q mutation and the CC1 deletion resulted in a twofold increase in SMAC ubiquitination, suggesting that the tight, multivalent tethering of SMAC also functions to protect it from becoming ubiquitinated. We observed no further increase in ubiquitination for the variant mutated for all three sites, presumably because of the overall weakening of the SMAC:BIRC6 interaction. In sum, these data suggest that the three characterized SMAC interaction sites are critical to limit its ubiquitination and allow for robust inhibition of BIRC6.

Discussion

IAPs function as cellular safeguards controlling programmed cell death. To achieve this, IAPs must ensure that caspases are kept inactive under nonstressed conditions while allowing their SMAC-induced release during apoptotic stimuli. To address the molecular mechanism underlying this fundamentally important decision on cell survival and cell death, we performed a mechanistic analysis of BIRC6, the only essential IAP. As revealed by its cryo-EM structure, the giant ubiquitin ligase adopts a C-shaped fold around a central cavity containing various receptor sites. The cavity fosters competition among substrates, with SMAC using most of the provided contact points and thus outcompeting other clients. We thus propose that SMAC, upon release from damaged mitochondria, undergoes multivalent interactions with IAPs such as BIRC6 to free the sequestered caspases, which then complete the cell death program. Moreover, multivalent binding protects SMAC from being ubiquitinated itself, explaining how the inhibitor can persist when bound to BIRC6 and ensure completion of apoptosis. Conversely, under nonstressed conditions when SMAC is absent, BIRC6 can target and ubiquitinate N-degron proteases. This basal activity is critical to prevent the untimely apoptosis caused by stochastic leaking of pro-apoptotic factors from mitochondria (Fig. 3F) (*14*).

The regulation and activity of BIRC6 is tied closely to key cellular processes. Clearance of toxic aggregates, for example, benefits from BIRC6 knockdown, because elevated ATG8 levels bolster the autophagy response. By studying the ubiquitination of LC3B, we show that BIRC6 uses an overlapping tandem AIM to target ATG8 proteins. These motifs do not coincide with the binding epitopes of HTRA2 and other apoptotic proteases, allowing a separate regulation of cell death and cell survival pathways. SMAC inhibits ubiquitination of both apoptotic factors and LC3B. We reason that SMAC serves as a general regulator to switch BIRC6 activity on and off in response to stress stimuli requiring coregulation of autophagy and apoptosis pathways (*31*, *32*).

Our structural data pave the way for functional contextualization of known posttranslational modifications and disease mutations within BIRC6 (Fig. 3F). For instance, an Arg3204Gln mutation in mice has been linked with developmental disorders (*33*, *34*). The corresponding Arg3190 in humans is part of the CBD-3 arginine loop that interacts with the negatively charged surfaces of SMAC and HTRA2. Accordingly, the mutation may influence the substrate spectrum of BIRC6 or, alternatively, it may affect the binding mode and thus functional state of the pro-apoptotic protease HTRA2. Further, proteomics studies identified ubiquitination of Lys2270 within the SMAC interface of CC1 as the most abundant posttranslational modification, possibly altering BIRC6 function (*35*). A regulatory role of CC1 is consistent with our biochemical data showing that loss of CC1 correlates to decreased inhibition by SMAC. Consistent with this observation, CC1 residues in the SMAC interface are the most frequently substituted BIRC6 residues detected in cancer (*36*). The most mutated site, Arg2291, is one of several conserved basic residues interacting with the negatively charged SMAC molecule. We hypothesize that CC1 mutations hinder SMAC binding and prevent apoptosis initiation. Tumors often present elevated levels of IAPs as a broad survival mechanism. Mutations disrupting the tripartite interface may thus desensitize cells to SMAC-triggered apoptosis without impairing other crucial functions of BIRC6 for cell survival and tumor propagation. It appears that multivalent binding of SMAC offers an additional layer to regulate BIRC6 activity and balance mitochondrial stress, apoptosis, and autophagy.

REFERENCES AND NOTES

1. A. H. Wyllie, J. F. Kerr, A. R. Currie, *Int. Rev. Cytol.* **68**, 251–306 (1980).
2. J. C. Reed, *J. Clin. Oncol.* **17**, 2941–2953 (1999).
3. C. B. Thompson, *Science* **267**, 1456–1462 (1995).
4. Q. L. Deveraux, J. C. Reed, *Genes Dev.* **13**, 239–252 (1999).
5. C. Du, M. Fang, Y. Li, L. Li, X. Wang, *Cell* **102**, 33–42 (2000).
6. A. M. Verhagen *et al*., *Cell* **102**, 43–53 (2000).
7. Y. Yang, S. Fang, J. P. Jensen, A. M. Weissman, J. D. Ashwell, *Science* **288**, 874–877 (2000).
8. Z. Chen *et al*., *Biochem. Biophys. Res. Commun.* **264**, 847–854 (1999).
9. H. P. Hauser, M. Bardroff, G. Pyrowolakis, S. Jentsch, *J. Cell Biol.* **141**, 1415–1422 (1998).
10. A. T. Nguyen *et al*., *Science* **357**, eaan0218 (2017).
11. J. Schelpe, D. Monté, F. Dewitte, T. K. Sixma, P. Rucktooa, *J. Biol. Chem.* **291**, 630–639 (2016).
12. K. Yanagitani, S. Juszkiewicz, R. S. Hegde, *Science* **357**, 472–475 (2017).
13. Y. Hao *et al*., *Nat. Cell Biol.* **6**, 849–860 (2004).
14. T. Bartke, C. Pohl, G. Pyrowolakis, S. Jentsch, *Mol. Cell* **14**, 801–811 (2004).
15. L. Vande Walle *et al*., *J. Proteome Res.* **6**, 1006–1015 (2007).
16. K. Sekine *et al*., *Biochem. Biophys. Res. Commun.* **330**, 279–285 (2005).
17. P. Ebner *et al*., *Nat. Commun.* **9**, 599 (2018).

18. C. Ge *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E1210–E1219 (2015).
19. L. Che *et al.*, *PLOS ONE* **14**, e0216553 (2019).
20. C. Ge *et al.*, *Hepatology* **69**, 2608–2622 (2019).
21. R. Jia, J. S. Bonifacino, *eLife* **8**, e50034 (2019).
22. S. Martens, D. Fracchiolla, *Cell Discov.* **6**, 23 (2020).
23. F. Mueller *et al.*, *Sci. Adv.* **7**, eabc8590 (2021).
24. P. C. A. da Fonseca *et al.*, *Nature* **470**, 274–278 (2011).
25. W. Li *et al.*, *Nat. Struct. Biol.* **9**, 436–441 (2002).
26. Y. Toyama, R. W. Harkness, T. Y. T. Lee, J. T. Maynes, L. E. Kay, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2025022118 (2021).
27. A. A. Jeyaprakash, C. Basquin, U. Jayachandran, E. Conti, *Structure* **19**, 1625–1634 (2011).
28. N. N. Noda, Y. Ohsumi, F. Inagaki, *FEBS Lett.* **584**, 1379–1385 (2010).
29. Z. Liu *et al.*, *Nature* **408**, 1004–1008 (2000).
30. E. Mastrangelo *et al.*, *Biophys. J.* **108**, 714–723 (2015).
31. G. Mariño, M. Niso-Santano, E. H. Baehrecke, G. Kroemer, *Nat. Rev. Mol. Cell Biol.* **15**, 81–94 (2014).
32. M. C. Maiuri, E. Zalckvar, A. Kimchi, G. Kroemer, *Nat. Rev. Mol. Cell Biol.* **8**, 741–752 (2007).
33. A. Enikanolaiye *et al.*, *Genome Res.* **30**, 540–552 (2020).
34. Z. A. Klugman, "Uncovering a novel role for BIRC6 in a dwarf mutant mouse," thesis, University of Toronto, Toronto, Ontario, Canada (2021).
35. P. V. Hornbeck *et al.*, *Nucleic Acids Res.* **43** (D1), D512–D520 (2015).
36. J. G. Tate *et al.*, *Nucleic Acids Res.* **47** (D1), D941–D947 (2019).

ACKNOWLEDGMENTS

We thank the staff of the Vienna BioCenter Core Facilities for their support, especially H. Kotisch and T. Heuser (Electron Microscopy), J. Neuhold (Protein Technologies), and the Mass Spectrometry Department; the EM facility at the Institute of Science and Technology Austria (ISTA), especially V.-V. Hodirnau, for measurement time and support; Life Science Editors for editing support; all members of the Clausen laboratory for discussions and J. Ahel for computational support; members of the Dagdas laboratory (Gregor Mendel Institute) for gifting the AIM peptides and LC3B plasmid; members of the Ehrmann laboratory (University Duisburg-Essen) for gifting HTRA2 plasmids; and our colleagues F. Ikeda (Osaka University, Japan) and Y. Dagdas (Gregor Mendel Institute, Vienna) for providing feedback on this study. **Funding:** This work was supported by Austrian Science Fund DocFunds grant DOC 112-B (J.F.E.); Marie Skłodowska-Curie grant 847548 (D.B.G.); Austrian Science Fund grant SFB F 79 (R.G.); and Austrian Research Promotion Agency Headquarter grant 852936. The IMP is supported by Boehringer Ingelheim. **Author contributions:** J.F.E., M.H., L.D., R.K., A.S., and R.G. prepared expression constructs and performed protein purifications. J.F.E. and D.B.G. performed biochemical experiments. J.F.E. prepared and screened cryo-EM specimens and performed initial processing. D.B.G. performed cryo-EM processing and model building. O.H. performed cross-link coupled mass spectrometry analysis. A.M. and G.V. helped to analyze data. J.F.E. and D.B.G. designed experiments. T.C. outlined the study. J.F.E., D.B.G., and T.C. prepared the manuscript with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Cryo-EM maps and atomic coordinates have been deposited in the Protein Data Bank (PDB) with accession codes 8ATU, 8ATX, 8AUK, and 8AUW and the Electron Microscopy Data Bank (EMDB) with accession codes EMD-15663, EMD-15668, EMD-15672, and EMD-15675. Uncropped and Coomassie-stained gels are provided in fig. S10, and XL-MS data are provided as data S1.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade8873
Materials and Methods
Figs. S1 to S10
Tables S1 and S2
References (*37*–*48*)
Data S1
MDAR Reproducibility Checklist

Submitted 14 September 2022; accepted 30 January 2023
10.1126/science.ade8873

STRUCTURE PREDICTION

Evolutionary-scale prediction of atomic-level protein structure with a language model

Zeming Lin[1,2]†, Halil Akin[1]†, Roshan Rao[1]†, Brian Hie[1,3]†, Zhongkai Zhu[1], Wenting Lu[1], Nikita Smetanin[1], Robert Verkuil[1], Ori Kabeli[1], Yaniv Shmueli[1], Allan dos Santos Costa[4], Maryam Fazel-Zarandi[1], Tom Sercu[1], Salvatore Candido[1], Alexander Rives[1,2]*

Recent advances in machine learning have leveraged evolutionary information in multiple sequence alignments to predict protein structure. We demonstrate direct inference of full atomic-level protein structure from primary sequence using a large language model. As language models of protein sequences are scaled up to 15 billion parameters, an atomic-resolution picture of protein structure emerges in the learned representations. This results in an order-of-magnitude acceleration of high-resolution structure prediction, which enables large-scale structural characterization of metagenomic proteins. We apply this capability to construct the ESM Metagenomic Atlas by predicting structures for >617 million metagenomic protein sequences, including >225 million that are predicted with high confidence, which gives a view into the vast breadth and diversity of natural proteins.

The sequences of proteins at the scale of evolution contain an image of biological structure and function. The biological properties of a protein constrain the mutations to its sequence that are selected through evolution, recording biology into evolutionary patterns (*1*–*3*). Protein structure and function can therefore be inferred from the patterns in sequences (*4*, *5*). This insight has been central to progress in computational structure prediction starting from classical methods (*6*, *7*) through the introduction of deep learning (*8*–*11*) up to present high-accuracy structure prediction (*12*, *13*).

Language models have the potential to learn patterns in protein sequences across evolution. This idea motivates research on evolutionary-scale language models (*14*), in which basic models (*15*–*17*) learn representations that reflect aspects of the underlying biology and, with greater representational capacity, capture secondary structure (*14*, *18*) and tertiary structure (*14*, *19*–*21*) at a low resolution.

Beginning with Shannon's model for the entropy of text (*22*), language models of increasing complexity have been developed, which has culminated in modern large-scale attention-based architectures (*23*–*25*). Despite the simplicity of their training objectives, such as filling in missing words or predicting the next word, language models of text are shown to exhibit emergent capabilities that develop as a function of scale in increasing computational power, data, and number of parameters. Modern language models containing tens to hundreds of billions of parameters show abilities such as few-shot language translation, commonsense reasoning, and mathematical problem solving, all without explicit supervision (*26*–*29*).

We posit that the task of filling in missing amino acids in protein sequences across evolution will require a language model to understand the underlying structure that creates the patterns in the sequences. As the representational capacity of the language model and the diversity of protein sequences seen in its training increase, we expect deep information about the biological properties of the protein sequences to emerge because those properties give rise to the patterns that are observed in the sequences. To study this kind of emergence, we scale language models from 8 million parameters up to 15 billion parameters. We discover that atomic-resolution structure emerges and continues to improve in language models over the four orders of magnitude in parameter scale. Strong correlations between the language model's understanding of the protein sequence (perplexity) and the accuracy of the structure prediction reveal a close link between language modeling and the learning of structure.

We show that language models enable fast end-to-end atomic-resolution structure prediction directly from sequence. Our approach leverages the evolutionary patterns that are captured by the language model to produce accurate atomic-level predictions. This removes costly aspects of the current state-of-the-art structure prediction pipeline, which eliminates the need for a multiple sequence alignment (MSA) while greatly simplifying the neural architecture used for inference. This results in an improvement in speed of up to 60× on the inference forward pass alone while also removing the search process for related proteins entirely, which can take >10 min with

[1]FAIR, Meta AI, New York, NY, USA. [2]New York University, New York, NY, USA. [3]Stanford University, Palo Alto, CA, USA. [4]Massachusetts Institute of Technology, Cambridge, MA, USA.
***Corresponding author. Email: arives@meta.com**
†These authors contributed equally to this work.

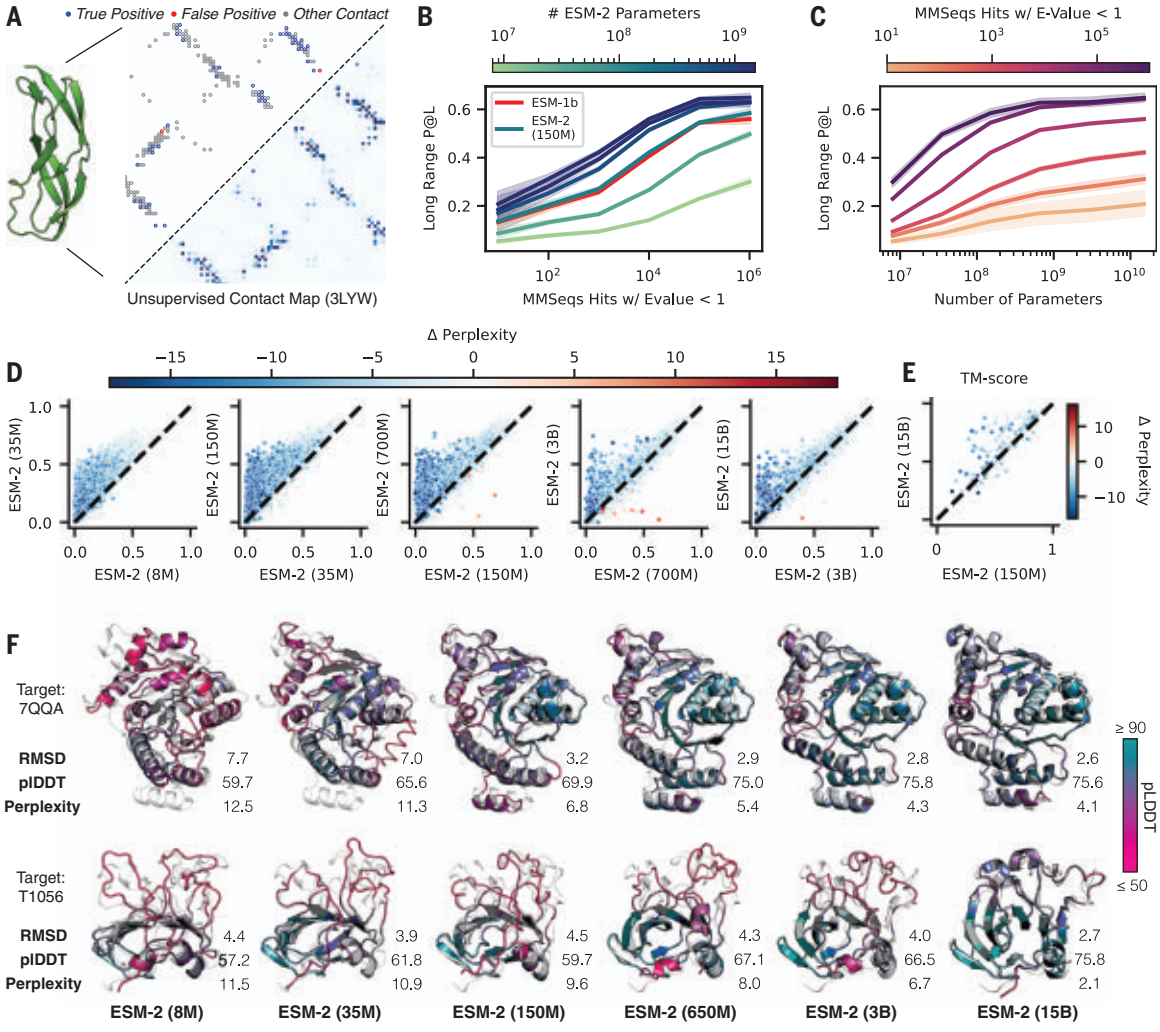


Fig. 1. Emergence of structure when scaling language models to 15 billion parameters. (**A**) Predicted contact probabilities (bottom right) and actual contact precision (top left) for PDB 3LYW. A contact is a positive prediction if it is within the top L most likely contacts for a sequence of length L. (**B** to **D**) Unsupervised contact prediction performance [long-range precision at L (P@L)] (SM A.2.1) for all scales of the ESM-2 model. (B) Performance binned by the number of MMseqs hits when searching the training set. Larger ESM-2 models perform better at all levels; the 150-million-parameter ESM-2 model is comparable to the 650-million-parameter ESM-1b model. (C) Trajectory of improvement as model scale increases for sequences with different numbers of MMseqs hits. (D) Left-to-right shows models from 8 million to 15 billion parameters, comparing the smaller model (*x* axis) against the next larger model (*y* axis) through unsupervised contact precision. Points are PDB proteins colored by change in perplexity for the sequence between the smaller and larger model. Sequences with large changes in contact prediction performance also exhibit large changes in language model understanding measured by perplexity. (**E**) TM-score on combined CASP14 and CAMEO test sets. Predictions are made by using structure module–only head on top of language models. Points are colored by the change in perplexity between the models. (**F**) Structure predictions on CAMEO structure 7QQA and CASP target 1056 at all ESM-2 model scales, colored by pLDDT (pink, low; teal, high). For 7QQA, prediction accuracy improves at the 150-million-parameter threshold. For T1056, prediction accuracy improves at the 15-billion-parameter threshold.

the high-sensitivity pipelines used by AlphaFold (*12*) and RoseTTAFold (*13*) and is a meaningful part of the computational cost even with recent lower-sensitivity fast pipelines (*30*). In practice, this means the speedup over the state-of-the-art prediction pipelines is up to one to two orders of magnitude.

This speed advantage makes it possible to expand structure prediction to metagenomic scale datasets. The past decade has seen efforts to expand knowledge of protein sequences to the immense microbial natural diversity of Earth through metagenomic sampling. These efforts have contributed to an exponential growth in the size of protein sequence databases, which now contain billions of proteins (*31*–*33*). Computational structural characterizations have recently been completed for ~20,000 proteins in the human proteome (*34*) and the ~200 million cataloged proteins of Uniprot (*35*), but the vast scale of metagenomic proteins represents a far greater challenge for structural characterization. The extent and diversity of metagenomic structures is unknown and is a frontier for biological knowledge, as well as a potential source of discoveries for medicine and biotechnology (*36*–*38*).

We present an evolutionary-scale structural characterization of metagenomic proteins that folds practically all sequences in MGnify90 (*32*), >617 million proteins. We were able to complete this characterization in 2 weeks on a heterogeneous cluster of 2000 graphics processing units (GPUs), which demonstrates scalability to far larger databases. High-confidence predictions are made for >225 million structures, which reveals and characterizes regions of metagenomic space distant from existing knowledge. Most (76.8%) high-confidence predictions are separate from UniRef90 (*39*) by at least 90% sequence identity, and tens of millions of predictions (12.6%) do not have any match to experimentally determined structures. These results give a large-scale view into the vast extent and diversity of metagenomic protein structures. These predicted structures can be accessed in the ESM Metagenomic Atlas (https://esmatlas.com) open science resource.

Atomic-resolution structure emerges in language models trained on protein sequences

We begin with a study of the emergence of high-resolution protein structure. We trained a family of transformer protein language models,

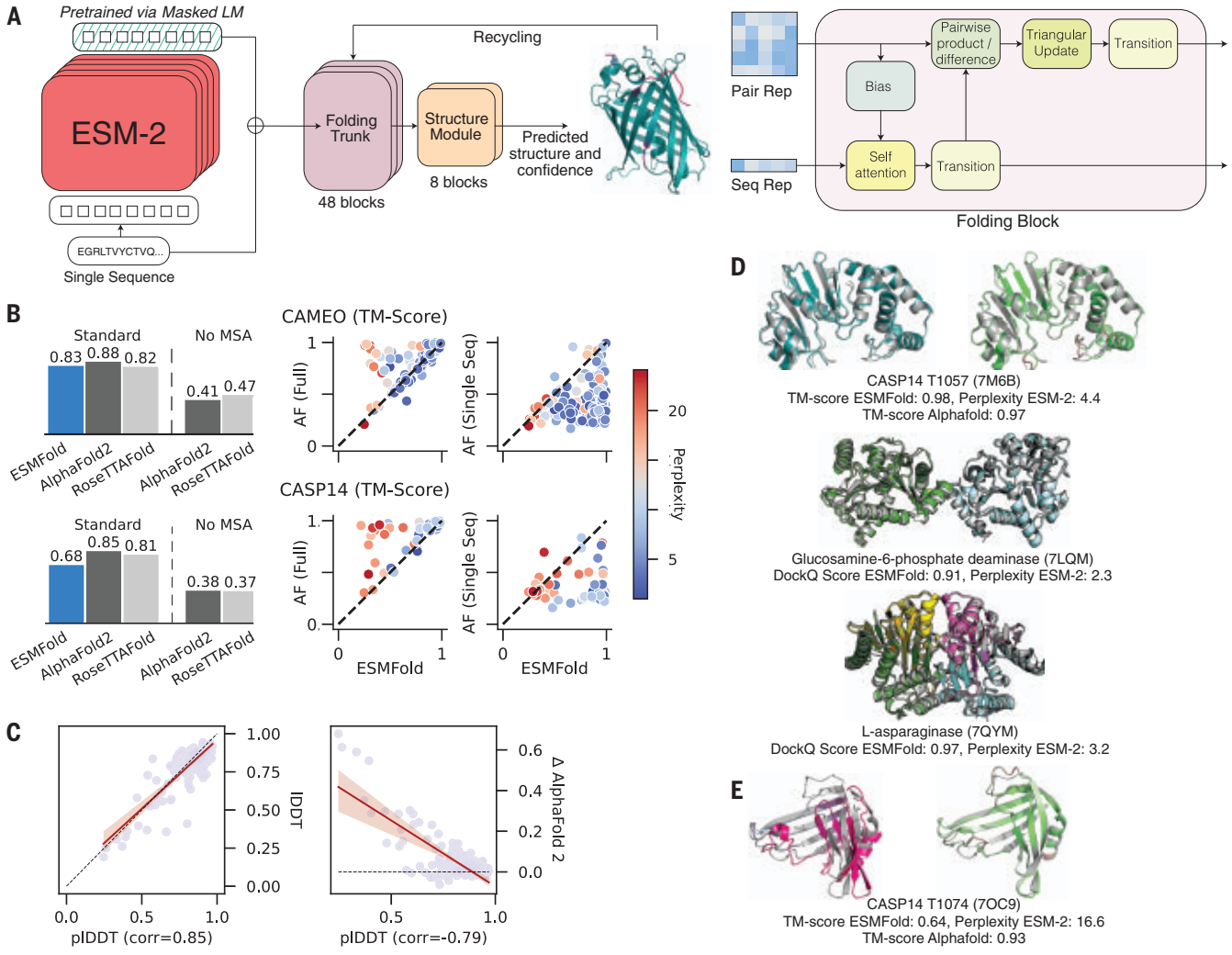


Fig. 2. Single sequence structure prediction with ESMFold. (**A**) ESMFold model architecture. Arrows show the information flow in the network from the language model to the folding trunk to the structure module that outputs 3D coordinates and confidences. LM, language model. (**B**) ESMFold produces accurate atomic resolution predictions, with similar accuracy to RoseTTAFold on CAMEO. When MSAs are ablated for AlphaFold and RoseTTAFold, performance of the models degrades. Scatterplots compare ESMFold (*x* axis) predictions with AlphaFold2 (*y* axis), colored by language model perplexity. Proteins with low perplexity score similarly to AlphaFold2. AF, AlphaFold2. (**C**) Model pLDDT versus true LDDT (left) and relative performance against AlphaFold (right) on CAMEO. pLDDT is a well-calibrated estimate of prediction accuracy. (**D**) Successful examples: Top shows test-set predictions of T1057, with ESMFold (left) and AlphaFold2 (right). Coloring shows predicted LDDT for both models (ESMFold high confidence, teal; AlphaFold2 high confidence, green; both low confidence, pink). Ground truth is shown in gray. The bottom two show complex predictions on a dimer (PDB: 7LQM) and a tetramer (PDB: 7QYM); ESMFold predictions are colored by chain ID and overlaid on ground truth (gray). DockQ (*50*) scores are reported for the interactions; in the case of the tetramer 7QYM, the score is the average of scores over interacting chain pairs. (**E**) Unsuccessful example: test-set predictions of T1074, with ESMFold (left) and AlphaFold2 (right). Coloring shows predicted LDDT for both models (ESMFold high confidence, teal; AlphaFold2 high confidence, green; both low confidence, pink). Ground truth is shown in gray. ESMFold TM-score is substantially below AlphaFold2 TM-score. The perplexity of the unsuccessful sequence is 16.6, meaning the language model does not understand the input sequence.

ESM-2, at scales from 8 million parameters up to 15 billion parameters. Relative to our previous generation model ESM-1b, ESM-2 introduces improvements in architecture, training parameters, and increases computational resources and data [supplementary material (SM) sections A.1.1 and A.2]. The resulting ESM-2 model family outperforms previously state-of-the-art ESM-1b (a ~650 million parameter model) at a comparable number of parameters, and on structure prediction benchmarks it also outperforms other recent protein language models (table S1).

ESM-2 is trained to predict the identity of amino acids that have been randomly masked out of protein sequences:

$$\mathcal{L}_{\text{MLM}} = -\sum_{i \in M} \log p(x_i | x_{\setminus M}) \quad (1)$$

where for a randomly generated mask M that includes 15% of positions i in the sequence x, the model is tasked with predicting the identity of the amino acids x_i in the mask from the surrounding context $x_{\setminus M}$, excluding the masked positions. This masked language modeling objective (*25*) causes the model to learn dependencies between the amino acids. Although the training objective itself is simple and unsupervised, solving it over millions of evolutionarily diverse protein sequences requires the model to internalize sequence patterns across evolution. We expect that this training will cause biological structure to

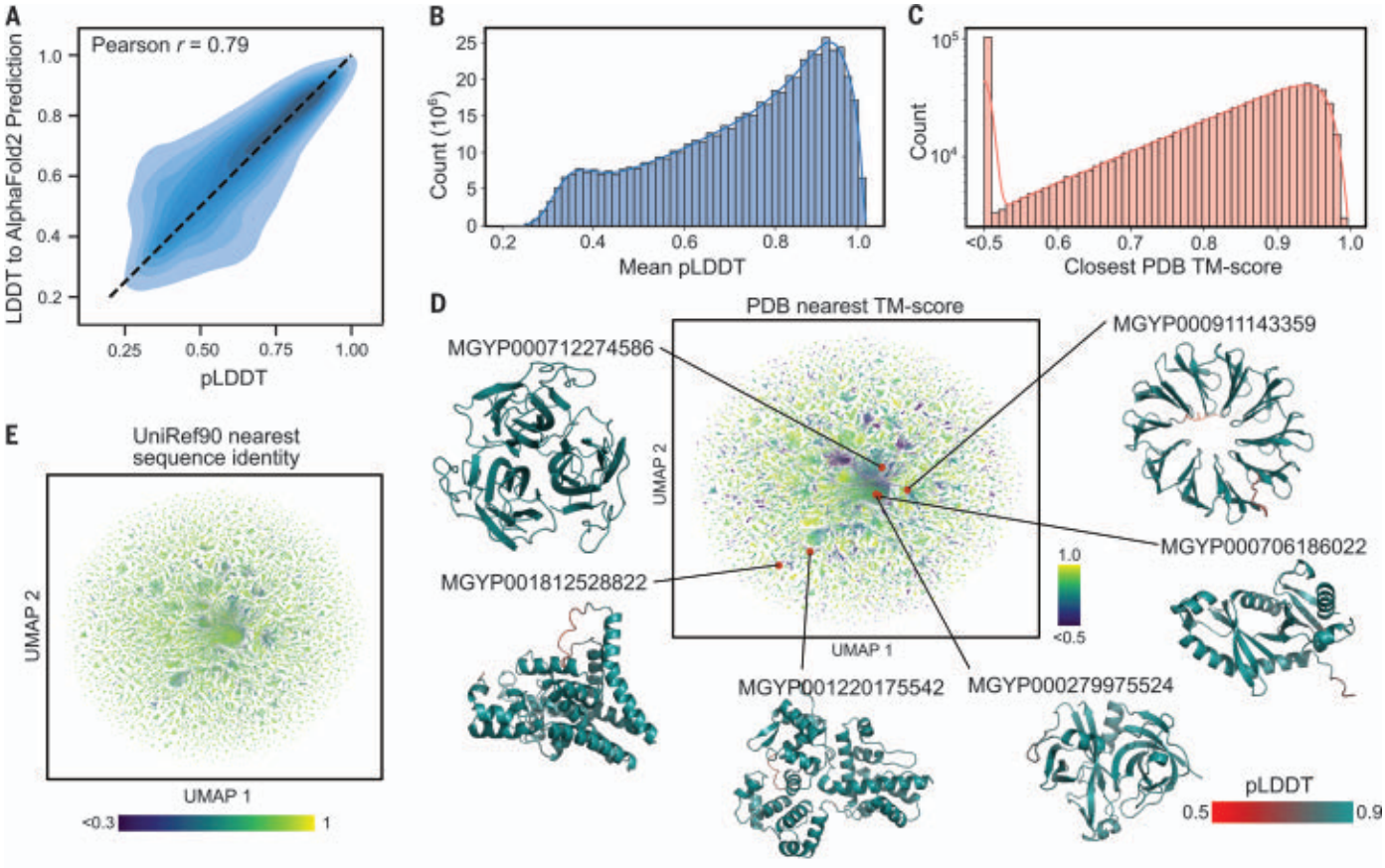


Fig. 3. Mapping metagenomic structural space. (**A**) ESMFold calibration with AlphaFold2 for metagenomic sequences. Mean pLDDT is shown on the *x* axis, and LDDT to the corresponding AlphaFold2 prediction is shown on the *y* axis. Distribution is shown as a density estimate across a subsample of ~4000 sequences from the MGnify database. (**B**) Distribution of mean pLDDT values computed for each of ~617 million ESMFold-predicted structures from the MGnify database. (**C**) The distribution of the TM-score to the most similar PDB structure for each of 1 million randomly sampled high-confidence (mean pLDDT > 0.7 and pTM > 0.7) structures. Values were obtained by a Foldseek search, which does not report values under 0.5 TM-score (*53*). (**D**) Sample of 1 million high-confidence protein structures is visualized in two dimensions by using the UMAP algorithm and colored according to distance from the nearest PDB structure, in which regions with low similarity to known structures are colored in dark blue. Example protein structures and their locations within the sequence landscape are provided; see also Fig. 4 and table S2. (**E**) Additional UMAP plot in which the 1 million sequences are plotted according to the same coordinates as in (D) but colored by the sequence identity to the most similar entry in UniRef90 according to a blastp (*60*) search.

materialize in the language model because it is linked to the sequence patterns. ESM-2 is trained over sequences in the UniRef (*39*) protein sequence database. During training, sequences are sampled with even weighting across ~43 million UniRef50 training clusters from ~138 million UniRef90 sequences, so that over the course of training, the model sees ~65 million unique sequences.

As we increase the scale of ESM-2 from 8 million to 15 billion parameters, we observe large improvements in the fidelity of its modeling of protein sequences. This fidelity can be measured by using perplexity, which ranges from 1 for a perfect model to 20 for a model that makes predictions at random. Intuitively, the perplexity describes the average number of amino acids that the model is choosing among for each position in the sequence. Mathematically, perplexity is defined as the exponential of the negative log-likelihood of the sequence (SM A.2.2). Figure S1 shows perplexity for the ESM-2 family as a function of the number of training updates, evaluated on a set of ~500,000 UniRef50 clusters that have been held out from training. Comparisons are performed at 270,000 training steps for all models in this section. The fidelity continues to improve as the parameters increase up to the largest model. The 8-million-parameter model has a perplexity of 10.45, and the 15 billion model reaches a perplexity of 6.37, which indicates a large improvement in the understanding of protein sequences with scale.

This training also results in the emergence of structure in the models. Because ESM-2's training is only on sequences, any information about structure that develops must be the result of representing the patterns in sequences. Transformer models that are trained with masked language modeling are known to develop attention patterns that correspond to the residue–residue contact map of the protein (*19*, *20*). We examine how this low-resolution picture of protein structure emerges as a function of scale. We use a linear projection to extract the contact map from the attention patterns of the language model (SM A.2.1). The precision of the top L (length of the protein) predicted contacts (long-range contact precision) measures the correspondence of the attention pattern with the structure of the protein. Attention patterns develop in ESM-2 that correspond to tertiary structure (Fig. 1A), and scaling leads to large improvements in the understanding of structure (Fig. 1B). The accuracy of the predicted contacts varies as a function of the number of evolutionarily related sequences in the training set. Proteins with more related sequences in the training set have steeper learning trajectories with respect to model scale (Fig. 1C). Improvement on sequences with high evolutionary

depth thus saturates at lower model scales, and improvement on sequences with low evolutionary depth continues as models increase in size.

For individual proteins, we often observe nonlinear improvements in the accuracy of the contact prediction as a function of scale. Plotting the change in the distribution of long-range contact precision at each transition to a higher level of scale reveals an overall shift in the distribution toward better performance (Fig. 1D), as well as a subset of proteins that undergo greater improvement. The accuracy of the contact map prediction and perplexity are linked, with proteins undergoing large changes in contact map accuracy also undergoing large changes in perplexity [normalized discounted cumulative gain (NDCG) = 0.87] (SM A.2.6). This link indicates that the language modeling objective is directly correlated with the materialization of the folded structure in the attention maps.

To identify atomic-resolution information in the model, we project out spatial coordinates for each of the atoms from the internal representations of the language model using an equivariant transformer (SM A.3.3). This projection is fitted by using experimentally determined protein structures from the Protein Data Bank (PDB) (*40*) and evaluated on 194 CAMEO proteins (*41*) and 51 CASP14 proteins (*42*). TM-score, which ranges from 0 to 1, measures the accuracy of the projection in comparison to the ground truth structure, with a value of 0.5 corresponding to the threshold for correctly predicting the fold (*43*). The evaluation uses a temporal cutoff, which ensures that the proteins used for testing are held out from those used in fitting the projection. This makes it possible to measure how atomic-level information emerges in the representations as a function of the parameter scale.

We discover that an atomic-resolution structure prediction can be projected from the representations of the ESM-2 language models. The accuracy of this projection improves with the scale of the language model. The 15 billion parameter model reaches a TM-score of 0.72 on the CAMEO test set and 0.55 on the CASP14 test set, a gain of 14 and 17% respectively relative to the 150 million parameter ESM-2 model (Fig. 1E). At each increase in scale a subset of proteins undergoes large changes in accuracy. For example, the protein 7QQA improves in root mean square deviation (RMSD) from 7.0 to 3.2 Å when the scale is increased from 35 million to 150 million parameters, and the CASP target T1056 improves in RMSD from 4.0 to 2.6 Å when the scale is increased from 3 billion to 15 billion parameters (Fig. 1F). Before and after these jumps, changes in RMSD are much smaller. Across all models (table S1), there is a correlation of −0.99 between validation perplexity and CASP14 TM-score and −1.00 between validation perplexity and CAMEO TM-score, which indicates a strong connection between the understanding of the sequence measured by perplexity and the atomic-resolution structure prediction. Additionally, there are strong correlations between the low-resolution picture of the structure that can be extracted from the attention maps and the atomic-resolution prediction (0.96 between long-range contact precision and CASP14 TM-score and 0.99 between long-range contact precision and CAMEO TM-score). These findings connect improvements in language modeling with the increases in low-resolution (contact map) and high-resolution (atomic-level) structural information.

Accelerating accurate atomic-resolution structure prediction with a language model

Language models greatly accelerate state-of-the-art high-resolution structure prediction. The language model internalizes evolutionary patterns linked to structure, which eliminates the need for external evolutionary databases, MSAs, and templates. We find that the ESM-2 language model generates state-of-the-art three-dimensional (3D) structure predictions directly from the primary protein sequence, which results in a speed improvement for structure prediction of more than an order of magnitude while maintaining high-resolution accuracy.

We developed ESMFold, a fully end-to-end single-sequence structure predictor, by training a folding head for ESM-2 (Fig. 2A). At prediction time, the sequence of a protein is inputted to ESM-2. The sequence is processed through the feedforward layers of the language model, and the model's internal states (representations) are passed to the folding head. The head begins with a series of folding blocks. Each folding block alternates between updating a sequence representation and a pairwise representation. The output of these blocks is passed to an equivariant transformer structure module, and three steps of recycling are performed before outputting a final atomic-level structure and predicted confidences (SM A.3.1). This architecture represents a major simplification in comparison with current state-of-the-art structure prediction models, which deeply integrate the MSA into the neural network architecture through an attention mechanism that operates across the rows and columns of the MSA (*12*, *44*).

Our approach results in a considerable improvement in prediction speed. On a single NVIDIA V100 GPU, ESMFold makes a prediction on a protein with 384 residues in 14.2 s, six times faster than a single AlphaFold2 model. On shorter sequences, the improvement increases up to ~60× (fig. S2). The search process for related sequences, which is required to construct the MSA, can take >10 min with the high-sensitivity protocols used by the published versions of AlphaFold and RoseTTAFold; this time can be reduced to <1 min, although with reduced sensitivity (*30*).

We train the folding head on ~25,000 clusters covering a total of ~325,000 experimentally determined structures from the PDB, which is further augmented with a dataset of ~12 million structures that we predicted with AlphaFold2 (SM A.1.2). The model is trained with the same losses that are used for AlphaFold (*45*). To evaluate the accuracy of structure predictions, we use test sets that are held out from the training data by a May 2020 cutoff date; as a result, all structures that are used in evaluation are held out from the training, and the evaluation is representative of the performance that would be expected in regular usage as a predictive model on the kinds of structures that are selected by experimentalists for characterization. This also makes it possible to compare with AlphaFold and RoseTTAFold because these models also have not been trained on structures deposited after May 2020. We use two test sets: The CAMEO test set consists of 194 structures that are used in the ongoing CAMEO assessment (between April 2022 and June 2022); the CASP14 test set consists of 51 publicly released structures that have been selected for their difficulty for the biannual structure prediction competition.

We compare the results of ESMFold on these evaluation sets to AlphaFold2 and RoseTTAFold (Fig. 2B). ESMFold achieves an average TM-score of 0.83 on CAMEO and 0.68 on CASP14. Using the search protocols released with AlphaFold2, including MSAs and templates, AlphaFold2 achieves 0.88 and 0.85 on CAMEO and CASP14, respectively. ESMFold achieves competitive accuracy with RoseTTAFold on CAMEO, which averages a TM-score of 0.82. When evaluating AlphaFold2 and RoseTTAFold on single sequences by ablating the MSA, their performance degrades substantially and falls well below that of ESMFold. This is an artificial setting because AlphaFold2 has not been explicitly trained for single sequences; however, it has recently emerged as important in protein design, in which these models have been used with single-sequence inputs for de novo protein design (*46*–*48*).

Although the average performance on the test sets is below AlphaFold2, the performance gaps are explained by the language model perplexity. On proteins for which perplexity is low, ESMFold results match AlphaFold2. On the CAMEO test set, the 3-billion-parameter ESM-2 model used in ESMFold achieves an average perplexity of 5.7. On the CASP14 test set, the same model only has an average perplexity of 10.0. Performance within each set

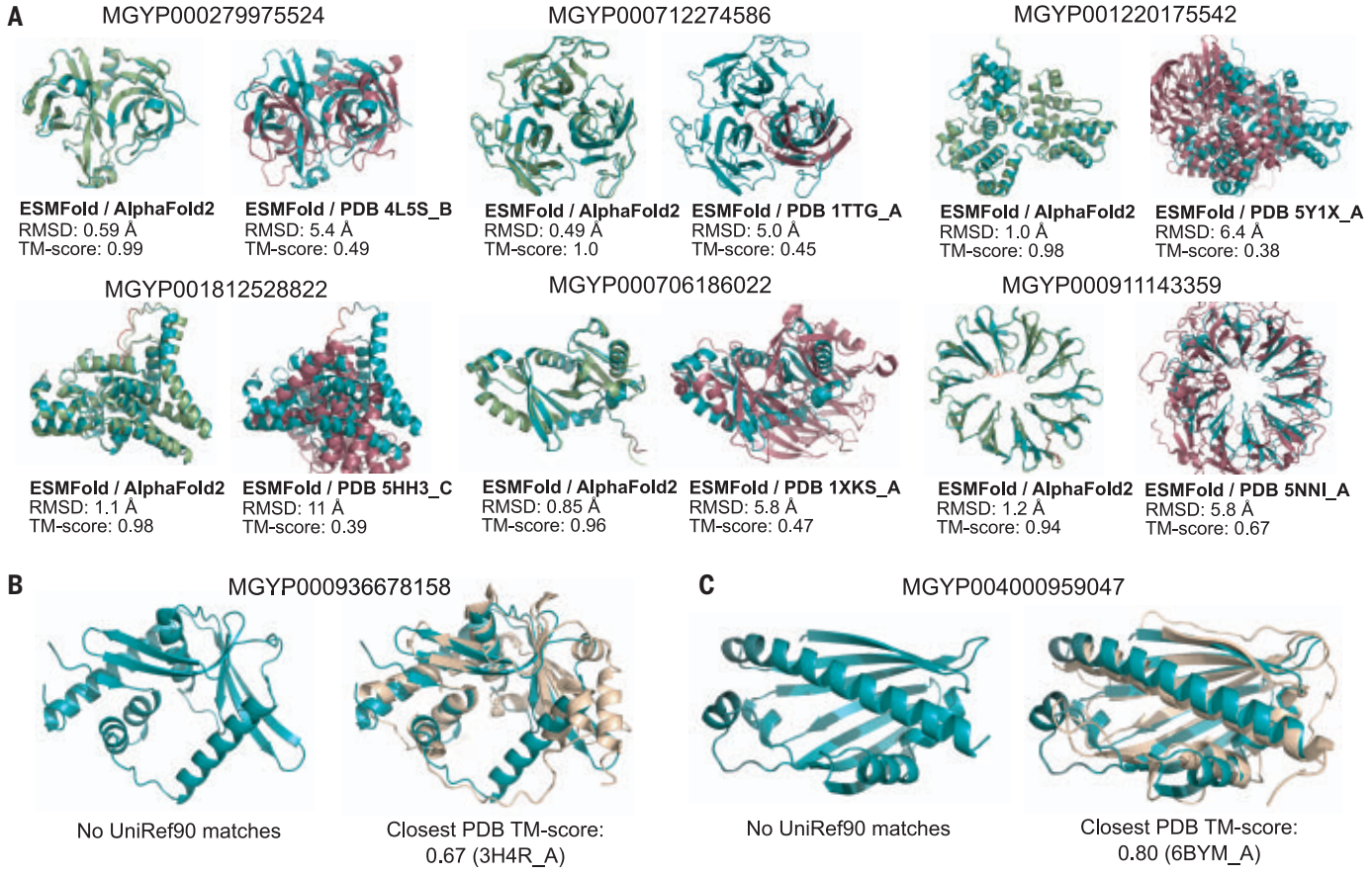


Fig. 4. Example ESMFold structure predictions of metagenomic sequences. (**A**) Example predicted structures from six different metagenomic sequences; also see table S2. Left of each subfigure: The prediction is displayed with the AlphaFold2 prediction (light green). Right of each subfigure: The prediction is displayed with the Foldseek-determined nearest PDB structure according to TM-score. (**B** and **C**) Examples of two ESMFold-predicted structures that have good agreement with experimental structures in the PDB but that have low sequence identity to any sequence in UniRef90. (B) Predicted structure of MGYP000936678158 aligns to an experimental structure from a bacterial nuclease (light brown, PDB: 3H4R), whereas (C) the predicted structure of MGYP004000959047 aligns to an experimental structure from a bacterial sterol binding domain (light brown, PDB: 6BYM).

is also well correlated with perplexity. On the CAMEO test set, language model perplexity has a Pearson correlation of −0.52 with the TM-score between the predicted and experimental structures; on CASP14, the correlation is −0.71 (Fig. 2B). On the subset of 18 CASP14 proteins for which ESM-2 achieves perplexity <7, ESMFold matches AlphaFold in performance (average TM-score difference <0.03 and no TM-score differences >0.1). The relationship between perplexity and structure prediction suggests that improvements in the language model will translate into improvements in single-sequence structure prediction accuracy, which is consistent with observations from the scaling analysis (Fig. 1, D and E). Additionally, this means that the language model's perplexity for a sequence can be used to predict the quality of the ESMFold structure prediction.

Ablation studies indicate that the language model representations are critical to ESMFold performance (fig. S3). With a folding trunk of eight blocks, performance on the CAMEO test set is 0.74 local distance difference test (LDDT) (baseline). Without the language model, this degrades substantially, to 0.58 LDDT. When removing the folding trunk entirely (i.e., only using the language model and the structure module), the performance degrades to 0.66 LDDT. Other ablations, such as only one block of a structure module, turning off recycling, not using AlphaFold2 predicted structures as distillation targets, or not using triangular updates, result in small performance degradations (change in LDDT of −0.01 to −0.04).

ESMFold provides state-of-the-art structure prediction accuracy, matching AlphaFold2 performance (<0.05 LDDT difference) on more than half the proteins (Fig. 2B). We find that this is true even on some large proteins—T1076 is an example with 0.98 TM-score and 540 residues (Fig. 2D). Parts of the structure with low accuracy do not differ notably between ESMFold and AlphaFold, which suggests that language models are learning information similar to that contained in MSAs. We also observe that ESMFold is able to make good predictions for components of homo- and heterodimeric protein-protein complexes (Fig. 2D). In a comparison with AlphaFold-Multimer (*49*) on a dataset of 2,978 recent multimeric complexes deposited in the PDB, ESMFold achieves the same qualitative DockQ (*50*) categorization for 53.2% of chain pairs, despite not being trained on protein complexes (fig. S4).

Confidence is well calibrated with accuracy. ESMFold reports confidence in the form of predicted LDDT (pLDDT) and predicted TM (pTM). This confidence correlates well with the accuracy of the prediction, and for high-confidence predictions (pLDDT > 0.7), the accuracy is comparable to AlphaFold2 (ESMFold LDDT = 0.83, AlphaFold2 LDDT = 0.85 on CAMEO) (Fig. 2C and fig. S5). High-confidence predictions approach experimental-level accuracy. On the CAMEO test set, ESMFold predictions have a median all-atom $RMSD_{95}$ (RMSD at 95% residue coverage) of 1.91 Å and backbone $RMSD_{95}$ of 1.33 Å. When confidence is very high (pLDDT > 0.9), predictions have median all-atom $RMSD_{95}$ of 1.42 Å and backbone $RMSD_{95}$ of 0.94 Å. The confidence can thus be used to predict how

likely it is that a given structure prediction will match the true structure if it were to be experimentally determined.

Recent work has investigated the use of language models for the direct prediction of protein structure from sequence, without a learned full atomic-level structure projection, but the accuracy has not been competitive with the use of MSAs (*21*, *51*). An approach developed concurrently with ours that uses a similar attention-based processing of language model representations to output atomic coordinates also appears to show results that are MSAs (*52*).

Evolutionary-scale structural characterization of metagenomics

This fast and high-resolution structure prediction capability enables the large-scale structural characterization of metagenomic proteins. We fold >617 million sequences from the MGnify90 database (*32*). This is the entirety of the sequences of length 20 to 1024 and covers 99% of all the sequences in MGnify90. Overall, the characterization produces ~365 million predictions with good confidence (mean pLDDT > 0.5 and pTM > 0.5), which corresponds to ~59% of the database, and ~225 million predictions with high confidence (mean pLDDT > 0.7 and pTM > 0.7), which corresponds to ~36% of total structures folded (Fig. 3). We were able to complete the predictions in 2 weeks on a cluster of ~2000 GPUs (SM A.4.1).

For structure prediction at scale, it is critical to distinguish well-predicted proteins from those that are poorly predicted. In the previous section, we evaluated calibration against experimentally determined structures on held-out test sets and found that the model confidence is predictive of the agreement with experimentally determined structures. We also assess calibration against AlphaFold predictions on metagenomic proteins. On a random subset of ~4000 metagenomic sequences, there is a high correlation (Pearson r = 0.79) between ESMFold pLDDT and the LDDT to AlphaFold2 predictions (Fig. 3A). When combined with results on CAMEO showing that when confidence is very high (pLDDT > 0.9), ESMFold predictions often approach experimental accuracy, these findings mean that ESMFold's confidence scores provide a good indication of the agreement with experimental structures and with the predictions that can be obtained from AlphaFold2. Across the ~617 million predicted structures, ~113 million structures meet the very high-confidence threshold.

Many of the metagenomic structure predictions have high confidence (Fig. 3B) and are not represented in existing structure databases (Figs. 3, C to E). On a random sample of 1 million high-confidence structures, 76.8% (767,580) of the proteins have a sequence identity below 90% to any sequence in UniRef90, which indicates that these proteins are distinct from existing UniRef90 sequences (Fig. 3E). For 3.4% (33,521 proteins), no match is found in UniRef90 at all (SM A.4.2). We use Foldseek (*53*) to compare the predicted structures with known structures in the PDB. At thresholds of 0.7 and 0.5 TM-score, Foldseek reports 25.4% (253,905 proteins) and 12.6% (125,765 proteins) without a match, respectively (Fig. 3, C and D). For 2.6% (25,664) there is both low structural similarity (TM-score ≤0.5) and no close sequence homolog (>30% identity). On the basis of these subsampled estimates, there are ~28 million proteins (12.6% of 225 million) with both high-confidence predictions and TM-score < 0.5 to known protein structures (examples in Fig. 4A and table S2). These results demonstrate that ESMFold can effectively characterize regions of protein space that are distant from existing knowledge.

Large-scale structural characterization also makes it possible to identify structural similarities in the absence of sequence similarity. Many high-confidence structures with low similarity to UniRef90 sequences do have similar structures in the PDB. This remote homology often extends beyond the limit detectable by sequence similarity. For example, MGnify sequence MGYP000936678158 has no matches to any entry in UniRef90 or through a jackhmmer (*54*) reference proteome search, but it has a predicted structure conserved across many nucleases (PDB: 5YET_B, TM-score 0.68; PDB: 3HR4_A, TM-score 0.67) (Fig. 4B and table S2); similarly, MGnify sequence MGYP004000959047 has no UniRef90 or jackhmmer reference proteome matches, but its predicted structure has a high similarity to the experimental structures of lipid binding domains (PDB: 6BYM_A, TM-score 0.80; PDB: 5YQP_B, TM-score 0.78) (Fig. 4C and table S2). The ability to detect remote similarities in structure enables insight into function that cannot be obtained from the sequence.

All predicted structures are available in the ESM Metagenomic Atlas (https://esmatlas.com) as an open science resource. Structures are available for bulk download, by means of an application programming interface (API), and through a web resource that provides search by structure and by sequence (*53*, *55*). These tools facilitate both large-scale and focused analysis of the full scope of the hundreds of millions of predicted structures.

Conclusions

Fast and accurate computational structure prediction has the potential to accelerate progress toward an era in which it is possible to understand the structure of all proteins discovered in gene sequencing experiments. Such tools promise insights into the vast natural diversity of proteins, most of which are discovered in metagenomic sequencing. To this end, we have completed a large-scale structural characterization of metagenomic proteins that reveals the predicted structures of hundreds of millions of proteins, millions of which are expected to be distinct in comparison to experimentally determined structures.

As structure prediction continues to scale to larger numbers of proteins, calibration becomes critical because, when the throughput of prediction is limiting, the accuracy and speed of the prediction form a joint frontier in the number of accurate predictions that can be generated. Very high-confidence predictions in the metagenomic atlas are expected to often be reliable at a resolution sufficient for insight similar to experimentally determined structures, such as into the biochemistry of active sites (*56*). For many more proteins for which the topology is predicted reliably, insight can be obtained into function through remote structural relationships that could not be otherwise detected with sequence.

The emergence of atomic-level structure in language models shows a high-resolution picture of protein structure encoded by evolution into protein sequences that can be captured with unsupervised learning. Our current models are very far from the limit of scale in parameters, sequence data, and computing power that can in principle be applied. We are optimistic that as we continue to scale, there will be further emergence. Our results showing the improvement in the modeling of low depth proteins point in this direction.

ESM-2 results in an advance in speed that in practical terms is up to one to two orders of magnitude, which puts far larger numbers of sequences within reach of accurate atomic-level prediction. Structure prediction at the scale of evolution can open a deep view into the natural diversity of proteins and accelerate the discovery of protein structures and functions.

REFERENCES AND NOTES

1. C. Yanofsky, V. Horn, D. Thorpe, *Science* **146**, 1593–1594 (1964).
2. D. Altschuh, T. Vernet, P. Berti, D. Moras, K. Nagai, *Protein Eng.* **2**, 193–199 (1988).
3. U. Göbel, C. Sander, R. Schneider, A. Valencia, *Proteins* **18**, 309–317 (1994).
4. A. S. Lapedes, B. G. Giraud, L. Liu, G. D. Stormo, *Lect. Notes Monogr. Ser.* **33**, 236–256 (1999).
5. J. Thomas, N. Ramakrishnan, C. Bailey-Kellogg, *IEEE/ACM Trans. Comput. Biol. Bioinform.* **5**, 183–197 (2008).
6. M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, T. Hwa, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 67–72 (2009).
7. F. Morcos *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, E1293–E1301 (2011).
8. S. Wang, S. Sun, Z. Li, R. Zhang, J. Xu, *PLOS Comput. Biol.* **13**, e1005324 (2017).
9. Y. Liu, P. Palmedo, Q. Ye, B. Berger, J. Peng, *Cell Syst.* **6**, 65–74.e3 (2018).
10. A. W. Senior *et al.*, *Nature* **577**, 706–710 (2020).
11. J. Yang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 1496–1503 (2020).
12. J. Jumper *et al.*, *Nature* **596**, 583–589 (2021).

13. M. Baek *et al*., *Science* **373**, 871–876 (2021).
14. A. Rives *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2016239118 (2021).
15. T. Bepler, B. Berger, *Cell Syst.* **12**, 654–669.e3 (2021).
16. E. C. Alley, G. Khimulya, S. Biswas, M. AlQuraishi, G. M. Church, *Nat. Methods* **16**, 1315–1322 (2019).
17. M. Heinzinger *et al*., *BMC Bioinformatics* **20**, 723 (2019).
18. A. Elnaggar *et al*., *IEEE Trans. Pattern Anal. Mach. Intell.* **14**, 1 (2021).
19. J. Vig *et al*., arXiv:2006.15222[cs, q-bio] (2021).
20. R. Rao, J. Meier, T. Sercu, S. Ovchinnikov, A. Rives, bioRxiv 422761 [Preprint] (2021); https://doi.org/10.1101/2020.12.15.422761.
21. R. Chowdhury *et al*., *Nat. Biotechnol.* **40**, 1617–1623 (2022).
22. C. E. Shannon, *Bell Syst. Tech. J.* **30**, 50–64 (1951).
23. A. Vaswani *et al*., in *Advances in Neural Information Processing Systems* (Curran Associates, 2017), pp. 5998–6008.
24. A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, Improving language understanding by generative pre-training (2018); https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.
25. J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (Association for Computational Linguistics, 2019), pp. 4171–4186.
26. T. B. Brown *et al*., in *Advances in Neural Information Processing Systems* (Curran Associates, 2020), pp. 1877–1901.
27. J. Wei *et al*., arXiv:2109.01652 [cs.CL] (2021).
28. J. Wei *et al*., arXiv:2201.11903 [cs] (2022).
29. A. Chowdhery *et al*., arXiv:2204.02311 [cs] (2022).
30. M. Mirdita *et al*., *Nat. Methods* **19**, 679–682 (2022).
31. M. Steinegger, M. Mirdita, J. Söding, *Nat. Methods* **16**, 603–606 (2019).
32. A. L. Mitchell *et al*., *Nucleic Acids Res.* **48**, D570–D578 (2020).
33. S. Mukherjee *et al*., *Nucleic Acids Res.* **49** (D1), D723–D733 (2021).
34. K. Tunyasuvunakool *et al*., *Nature* **596**, 590–596 (2021).
35. M. Varadi *et al*., *Nucleic Acids Res.* **50**, D439–D444 (2022).
36. O. Shimomura, F. H. Johnson, Y. Saiga, *J. Cell. Comp. Physiol.* **59**, 223–239 (1962).
37. K. Mullis *et al*., *Cold Spring Harb. Symp. Quant. Biol.* **51**, 263–273 (1986).
38. M. Jinek *et al*., *Science* **337**, 816–821 (2012).
39. B. E. Suzek, Y. Wang, H. Huang, P. B. McGarvey, C. H. Wu; UniProt Consortium, *Bioinformatics* **31**, 926–932 (2015).
40. S. K. Burley *et al*., *Nucleic Acids Res.* **47** (D1), D464–D474 (2019).
41. J. Haas *et al*., *Proteins* **86** (suppl. 1), 387–398 (2018).
42. A. Kryshtafovych, T. Schwede, M. Topf, K. Fidelis, J. Moult, *Proteins* **89**, 1607–1617 (2021).
43. Y. Zhang, J. Skolnick, *Proteins* **57**, 702–710 (2004).
44. R. M. Rao *et al*., in *Proceedings of the 38th International Conference on Machine Learning* (PMLR, 2021), pp. 8844–8856.
45. G. Ahdritz *et al*., bioRxiv 517210 [Preprint] (2022). https://doi.org/10.1101/2022.11.20.517210.
46. J. Dauparas *et al*., *Science* **378**, 49–56 (2022).
47. J. Wang *et al*., *Science* **377**, 387–394 (2022).
48. B. I. M. Wicky *et al*., *Science* **378**, 56–61 (2022).
49. R. Evans *et al*., bioRxiv 463034 [Preprint] (2021). https://doi.org/10.1101/2021.10.04.463034.
50. S. Basu, B. Wallner, *PLOS ONE* **11**, e0161879 (2016).
51. K. Weissenow, M. Heinzinger, B. Rost, *Structure* **30**, 1169–1177.e4 (2022).
52. R. Wu *et al*., bioRxiv 500999 [Preprint] (2022).
53. M. van Kempen *et al*., bioRxiv 479398 [Preprint] (2022). https://doi.org/10.1101/2022.02.07.479398.
54. S. C. Potter *et al*., *Nucleic Acids Res.* **46**, W200–W204 (2018).
55. M. Steinegger, J. Söding, *Nat. Biotechnol.* **35**, 1026–1028 (2017).
56. Y. Zhang, *Curr. Opin. Struct. Biol.* **19**, 145–155 (2009).
57. Z. Lin *et al*., ESM-2 and ESMFold-v0 Model Code and Weights, Zenodo (2023). https://doi.org/10.5281/zenodo.7566741.
58. Z. Lin *et al*., ESM Atlas v0 representative random sample of predicted protein structures, Zenodo (2022). https://doi.org/10.5281/zenodo.7623482.
59. Z. Lin *et al*., ESM Atlas v0 random sample of high confidence predicted protein structures, Zenodo (2022). https://doi.org/10.5281/zenodo.7623627.
60. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **215**, 403–410 (1990).

ACKNOWLEDGMENTS

We thank FAIR team members N. Goyal, Y. LeCun, A. Lerer, J. Liu, L. van der Maaten, and S. Sukhbaatar and collaborators J. Dauparas and S. Ovchinnikov for technical help, feedback, and discussions that helped shape this project. We thank E. Koonin and F. Zhang for feedback on the metagenomic dataset. We thank A. Rizvi, J. Shepard, and J. Spisak for program support. We thank S. Gomez, S. Jain, W. Ngan, and N. Seejoor for their work on the ESM Metagenomic Atlas website. We also thank the developers of the OpenFold project, fairseq, PyTorch, Foldseek, MMseqs2, PyMol, Biotite, and others for building invaluable open-source tools and the creators and maintainers of MGnify, PDB, UniProt, and UniRef, as well as the researchers whose experimental efforts are included in these resources. **Funding:** There were no external sources of funding for the project. **Author contributions:** Conceptualized and initiated the project: Z.L., S.C., and A.R.; developed and trained ESM-2: H.A., Z.Z., W.L., and R.V.; developed and trained ESMFold: Z.L., R.R., Z.Z., N.S., A.S.C., M.F.Z., and S.C.; produced metagenomic structure predictions: Z.L., H.A., and W.L.; analyzed ESM-2 and ESMFold: Z.L., R.R., Z.Z., and W.L.; analyzed predicted metagenomic structures: B.H. and T.S.; developed ESM Atlas: B.H., N.S., O.K., Y.S., and T.S.; wrote the manuscript: Z.L., H.A., R.R., B.H., and A.R.; engineering and science leadership: T.S., S.C., and A.R. **Competing interests:** The authors declare no competing financial interests. No patent applications have been filed on this work. **Data and materials availability:** All predicted structures in the ESM Metagenomic Atlas are available at https://esmatlas.com. ESM-2 and ESMFold model source code and parameters are available at https://github.com/facebookresearch/esm and archived on Zenodo (*57*). A representative random sample of ~1 million predicted structures is archived on Zenodo (*58*), and the random sample of ~1 million high-confidence predictions used for analysis in this work is also archived on Zenodo (*59*). Models and data have been released under permissive licenses (MIT license for model source code and parameters and CC-BY for predicted structures).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade2574
Materials and Methods
Supplementary Text
Figs. S1 to S8
Tables S1 to S5
References (*61*–*74*)

Submitted 3 August 2022; resubmitted 10 November 2022
Accepted 16 February 2023
10.1126/science.ade2574

NANOMATERIALS

Chemical scissor–mediated structural editing of layered transition metal carbides

Haoming Ding[1,2,3], Youbing Li[1,3], Mian Li[1,3], Ke Chen[1,3], Kun Liang[1,3], Guoxin Chen[3], Jun Lu[4], Justinas Palisaitis[4], Per O. Å. Persson[4], Per Eklund[4], Lars Hultman[4], Shiyu Du[1,2,3], Zhifang Chai[1,2,3], Yury Gogotsi[5]*, Qing Huang[1,3,6]*

Intercalated layered materials offer distinctive properties and serve as precursors for important two-dimensional (2D) materials. However, intercalation of non–van der Waals structures, which can expand the family of 2D materials, is difficult. We report a structural editing protocol for layered carbides (MAX phases) and their 2D derivatives (MXenes). Gap-opening and species-intercalating stages were respectively mediated by chemical scissors and intercalants, which created a large family of MAX phases with unconventional elements and structures, as well as MXenes with versatile terminals. The removal of terminals in MXenes with metal scissors and then the stitching of 2D carbide nanosheets with atom intercalation leads to the reconstruction of MAX phases and a family of metal-intercalated 2D carbides, both of which may drive advances in fields ranging from energy to printed electronics.

Intercalated materials are predominantly produced by introducing non-native species into the van der Waals (vdW) gaps of inherently layered vdW materials such as graphite, hexagonal boron nitride, and transition metal dichalcogenides (*1*, *2*). Guest-host interactions alter the electronic structure and enable property tailoring for energy storage, catalysis, electronic, optical, and magnetic applications (*3*–*7*). $M_{n+1}AX_n$, or "MAX phases," are a large family of ternary layered compounds that typically have weak metallic bonds between M and A atoms and covalent bonds within the $M_{n+1}X_n$ layers (*8*, *9*). Here, M denotes an early transition element, A is a main group element, X is nitrogen and/or carbon, and n is a number between one and four. The strong non-vdW bonding in MAX phases requires chemical etching of A elements to obtain their two-dimensional (2D) derivatives, MXenes (*10*–*12*). The resultant vdW gaps in MXenes provide space for intercalating various guest species. For example, anions such as F^-, O^{2-}, OH^-, and Cl^- spontaneously coordinate with exposed M atoms of MXenes as

[1]Engineering Laboratory of Advanced Energy Materials, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, Zhejiang 315201, China. [2]University of Chinese Academy of Sciences, 19 A Yuquan Rd, Shijingshan District, Beijing 100049, China. [3]Qianwan Institute of CNiTECH, Ningbo, Zhejiang 315336, China. [4]Thin Film Physics Division, Department of Physics, Chemistry, and Biology (IFM), Linköping University, SE-581 83 Linköping, Sweden. [5]Department of Materials Science and Engineering, A. J. Drexel Nanomaterials Institute, Drexel University, Philadelphia, PA 19104, USA. [6]Advanced Energy Science and Technology Guangdong Laboratory, Huizhou, Guangdong 516003, China.
***Corresponding authors. Email: huangqing@nimte.ac.cn (Q.H.); gogotsi@drexel.edu (Y.G.)**

termination species T, as described by the $M_{n+1}X_nT_x$ formula (*10*, *13*). Intercalation of cations, cationic surfactants, and organic molecules in vdW gaps expands the interlayer spacing of MXenes and facilitates their delamination into monolayers, finding roles in energy storage, printed electronics, electromagnetic interference shielding, and many other applications (*14*–*16*).

Recently, we reported a Lewis acidic molten salt (LAMS) etching protocol that is capable of both etching and substituting weakly bonded interlayer atoms in MAX phases (*13*). A series of MAX phases containing late transition metals and MXenes with pure halogen terminations were synthesized and explored for applications as catalysts and ferromagnetic and electrochemical energy storage materials (*17*–*20*). However, only a few LAMSs have thermophysical (solubility, melting point, and boiling point) and chemical (redox potential and activity of cations and coordination geometry of anions) properties required to act as both etchant and intercalant. For example, MXenes with -O, -S, -Se, -Te, and -NH terminations were only realized by an anion exchange reaction with brominated MXenes (*21*). The direct use of oxides or chalcogenides with strong covalent bonds to supply -O and chalcogen terminations would be a daunting task because of their high melting temperature and low solubility, both of which substantially limit the structural editing capability of LAMS etching. In this work, we introduce a chemical scissor–mediated intercalation chemistry for structural editing of non-vdW MAX phases and vdW MXenes. The range of constituent elements of MAX phases and terminating groups of MXenes is greatly extended. Structural editing by alternating LAMS and metal scissors leads to the exfoliation of MAX phases and MXenes into stacked lamellae directly in molten salts and guides the discovery of a series of 2D metal-intercalated layered carbides.

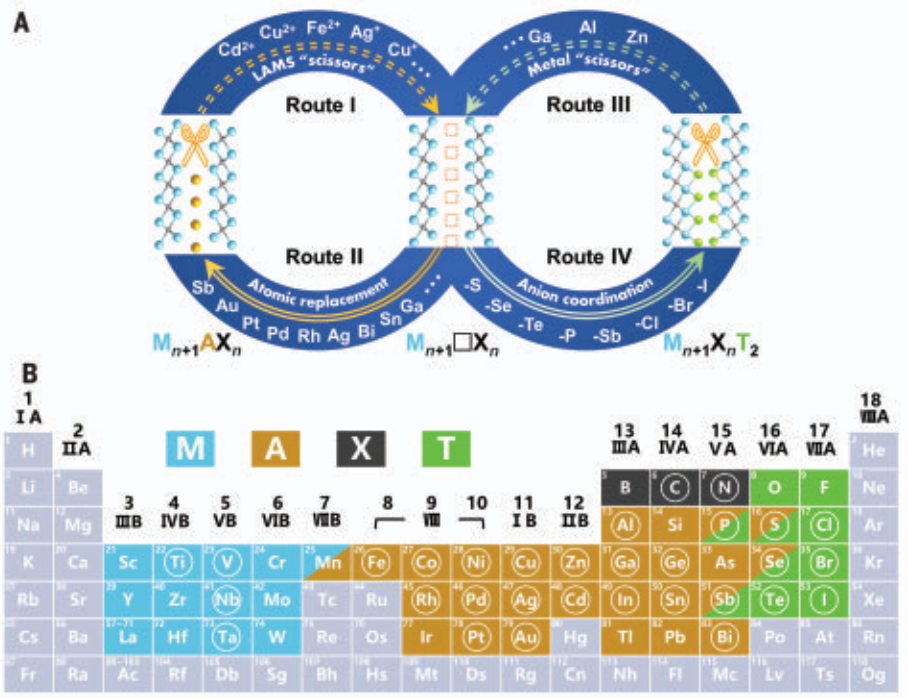


Fig. 1. Structural editing protocol of MAX and MXene mediated by the chemical scissors. (**A**) Schematic illustration of structural editing of MAX phases and MXenes through chemical scissor–mediated intercalation protocol. $M_{n+1}\square X_n$ denotes the structure with interlayer atom vacancies that formed after A-element etching. (**B**) Periodic table showing elements involved in the formation of MAX phases and MXenes. Light blue, M elements; brown, A elements; black, X elements; green, ligand (T) elements; circled, elements studied in the present work.

Chemical scissor–mediated structural editing routes

The chemical scissor–mediated structural editing protocol contains four reaction routes (Fig. 1A): (i) the opening of non-vdW gaps in MAX phases by LAMS scissors because of differences in redox potential between Lewis acidic cations and A elements (route I), (ii) the diffusion of metal atoms into interlayer atom vacancies to form MAX phases to lower the system's chemical energy (route II), (iii) the removal of surface terminations of multilayer MXenes through electron injection with metal scissors and opening of vdW gaps (route III), and (iv) the coordination of anions with oxidized early transition metal atoms to form terminated MXenes (route IV).

The periodic table emphasizes elements that are represented in MAX phases and MXenes (Fig. 1B). Aside from usual main-group elements (such as Al, Ga, and Sn), unconventional elements (Bi, Sb, Fe, Co, Ni, Cu, Zn, Pt, Au, Pd, Ag, Cd, and Rh) were intercalated into MAX phases. Meanwhile, in addition to the known halogen (-Cl, -Br, and -I) and chalcogen (-S, -Se, and -Te) terminations, the -P and -Sb (group 15 elements) terminations are demonstrated. All reaction recipes used in this study are listed in table S1.

Topotactic structural transformation of MAX phases aided by LAMS scissors

The Cu^{2+} cation in a LAMS scissor $CuCl_2$ has a strong electron affinity and can oxidize Al atoms that are weakly bonded in MAX phases (route I), as shown in Eqs. 1 and 2. As soon as interlayer atom vacancies V_A (denoted by □ in $M_{n+1}\square X_n$) are available, predissolved guest metal atoms A′ (e.g., Ga, In, or Sn) in molten salt diffuse into interlayers and occupy V_A to form $M_{n+1}A'X_n$ phases (Eq. 3) in a topotactic structural transformation manner (route II) (figs. S1 to S5). The A-element etching (vacancy formation) and intercalation (guest atom occupancy) are transient and concerted processes. The inserted guest atoms keep the interlayer space accessible to intercalants and prevent the etched $M_{n+1}\square X_n$ from collapsing into a close-packed, twin-like structure. Notably, the LAMS scissor should preferentially etch the A atoms in MAX phases but avoid the oxidation of intercalating metals (fig. S6) (*22*). Because of the thermodynamically favorable occupation of V_A vacancies, main-group metals with low melting points (T_m) diffuse into $M_{n+1}\square X_n$ to form stable MAX phases. Accordingly, with the aid of LAMS $CdCl_2$, Sb (T_m = 613°C) was successfully intercalated into a series of MAX phases such as Ti_3SbC_2, Ti_3SbCN, Nb_2SbC, and $Ti_3(Sb_{0.5}Sn_{0.5})CN$ (figs. S7 to S9). In the x-ray diffraction (XRD) pattern of Ti_3SbC_2, the (000*l*) peaks shifted toward higher Bragg angles compared with the Ti_3AlC_2 precursor (Fig. 2A), which indicates a shrinkage of lattice parameter *c* from 18.578 Å for Ti_3AlC_2 to 18.443 Å for Ti_3SbC_2 (fig. S10 and table S2). A similar decrease of lattice parameter *c*, but increase of *a*, was also observed in Nb_2SbC (*a* = 3.329 Å and *c* = 13.210 Å) as compared with Nb_2AlC (*a* = 3.107 Å and *c* = 13.888 Å). The strong coupling between the Sb 4p orbital with Ti 3d and Nb 4d orbitals accounts for the shortened bonding length along the *c* axis (*23*). Scanning transmission electron microscopy (STEM) imaging of Ti_3SbC_2 showed a typical zigzag pattern of the MAX phase along the $[11\bar{2}0]$ zone axis (Fig. 2B). Lattice-resolved

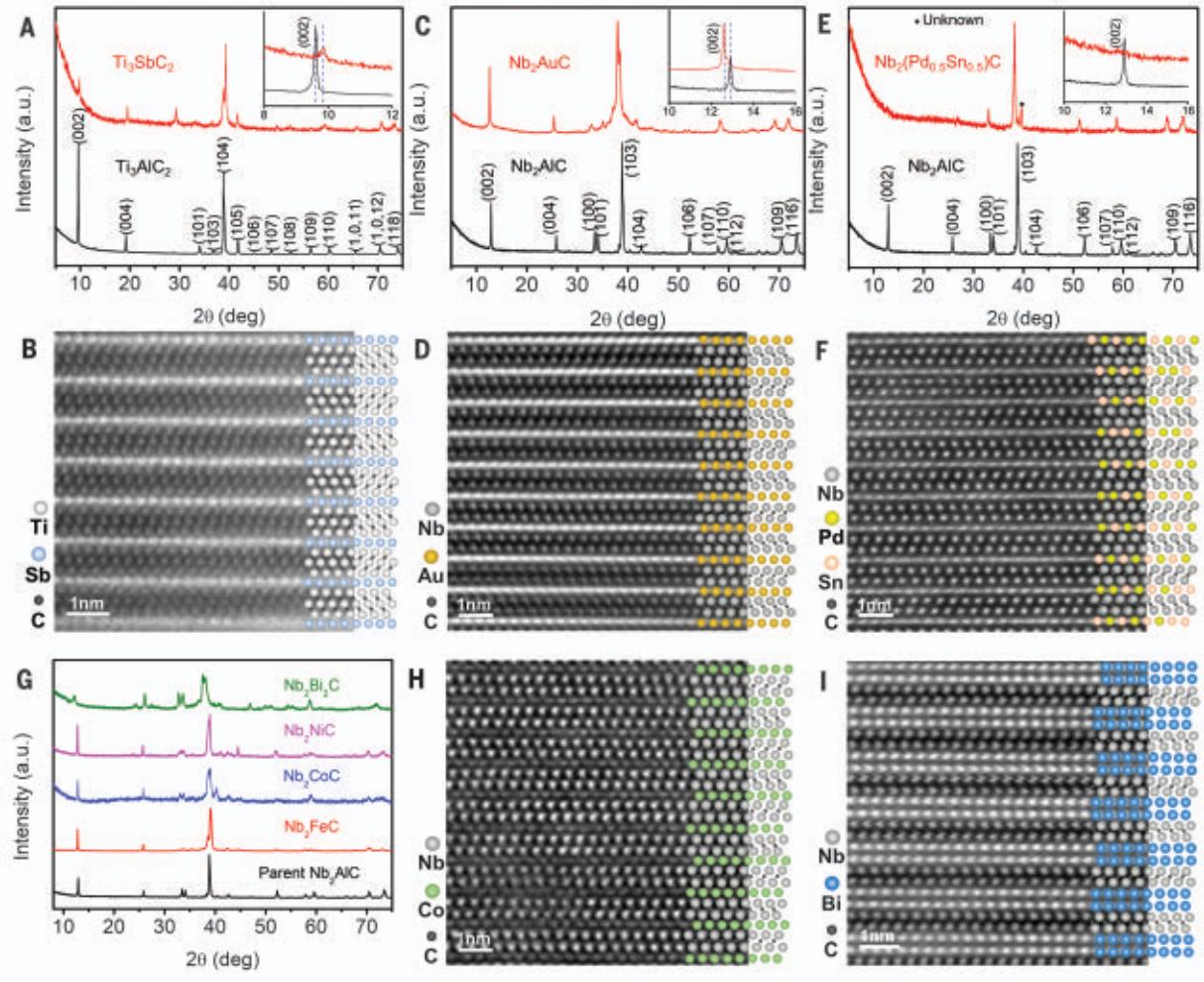


Fig. 2. Transformation of a MAX phase to another MAX phase. (**A**) XRD patterns of Ti_3SbC_2 and its parent phase Ti_3AlC_2. (**B**) STEM image of Ti_3SbC_2. (**C**) XRD patterns of Nb_2AuC and its parent phase Nb_2AlC. (**D**) STEM image of Nb_2AuC. (**E**) XRD patterns of $Nb_2(Pd_{0.5}Sn_{0.5})C$ and its parent phase Nb_2AlC. (**F**) STEM image of $Nb_2(Pd_{0.5}Sn_{0.5})C$. (**G**) XRD spectra of a series of MAX phases from Nb_2AlC: Nb_2FeC, Nb_2CoC, Nb_2NiC, and Nb_2Bi_2C. (**H** and **I**) STEM images of Nb_2CoC (H) and Nb_2Bi_2C (I). All STEM images were acquired along the $[11\bar{2}0]$ zone axis of MAX phases, and atomic structural models were added to corroborate the topotactic structural transition.

STEM combined with energy dispersive spectroscopy (EDS) and scanning electron microscopy (SEM) further corroborated the absence of Al in final Ti_3SbC_2, which indicates the complete substitution of Al by Sb through the LAMS scissor–mediated intercalation (fig. S11):

$$3Cu^{2+} + 2Al = 3Cu + 2Al^{3+} \quad (1)$$

$$M_{n+1}AlX_n = M_{n+1}\square X_n + Al^{3+} + 3e^- \quad (2)$$

$$M_{n+1}\square X_n + A' = M_{n+1}A'X_n \quad (3)$$

Noble metals are seldomly considered as constituent elements of MAX phases because of their chemical inertness and high melting points (*24*). However, low eutectic points (EPs) of noble metals alloys, such as Au-Cd (EP ≈ 629°C), Ag-Sb (EP ≈ 484°C), Pd-Sn (EP ≈ 600°C), Pt-Cd (EP ≈ 670°C), and Rh-Sn (EP ≈ 660°C) can promote the noble metal intercalation into the interlayer atom vacancy of $M_{n+1}\square X_n$ etched by LAMS scissor $CdCl_2$, and lead to the formation of noble metal–containing MAX phases: Nb_2AuC (Fig. 2, C and D, and fig. S12), $Nb_2(Au_{0.5}Al_{0.5})C$ (fig. S13), $Nb_2(Ag_{0.3}Sb_{0.4}Al_{0.3})C$ (fig. S14), $Nb_2(Pd_{0.5}Sn_{0.5})C$ (Fig. 2, E and F, and fig. S15), Nb_2PtC (fig. S16), $Nb_2(Pt_{0.6}Al_{0.4})C$ (fig. S17), and $Nb_2(Rh_{0.2}Sn_{0.4}Al_{0.4})C$ (fig. S18). Late transition metals in the fourth period (Fe, Co, Ni, Cu, and Zn) can also fully substitute for Al in Nb_2AlC aided by the same scissor, $CdCl_2$, and produce Nb_2FeC, Nb_2CoC, Nb_2NiC, Nb_2CuC, and Nb_2ZnC (Fig. 2, G and H, and figs. S19 to S24). The formation of MAX phases with transition metals in the A layer implicates the possibility of tuning their interlayer non-vdW bonding by orbital interaction of d-block electrons. However, the elements in groups 11 and 12 have saturated d orbitals, and their coupling with the d orbital of the transition metal becomes weak (*23*), which results in a small difference between the lattice parameters of Nb_2AuC (a = 3.175 Å and c = 14.062 Å) and Nb_2AlC (a = 3.107 Å and c = 13.888 Å), although the atomic radius of Au (146 pm) is much larger than that of Al (121 pm). The successful incorporation of noble metals with large atomic radii reflects the excellent structural tolerance and agile composition tunability in layered transition metal carbides.

In addition to monoatomic substitution, a double layer of Bi atoms in a MAX-like phase Nb_2Bi_2C was observed (Fig. 2I and fig. S25) in analogy with the well-studied Mo_2Ga_2C (*25*). This means that the chemical scissor–mediated protocol can not only enrich the elemental composition but can also expand the structural diversity of layered carbides.

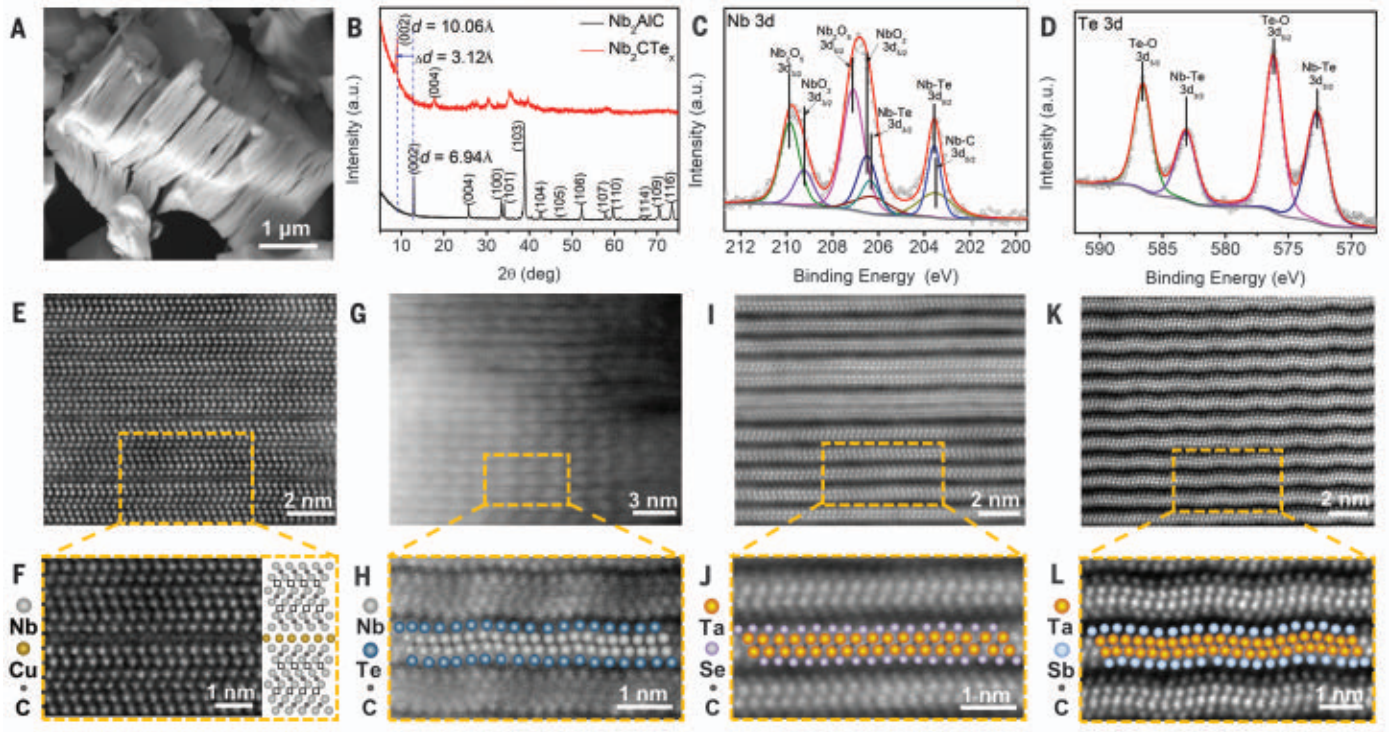


Fig. 3. Conversion from MAX phase to MXene. (**A**) SEM image of Nb_2CTe_x. (**B**) XRD patterns of Nb_2AlC and derived Nb_2CTe_x. Δd represents the swelling of interlayer spacing after chemical etching. (**C** and **D**) Nb 3d orbital (C) and Te 3d orbital (D) XPS spectra of Nb_2CTe_x. (**E**) STEM image revealing a close-packed, twin-like structure of Nb_2C and (**F**) the typical zigzag structure with interstitial voids denoted by black boxes in the structure, implicating the formation of interlayer atom vacancies after etched-out Al atoms. (**G**) STEM image of Nb_2CTe_x along the $[11\bar{2}0]$ zone axis and (**H**) their ripple-like morphology. (**I** and **J**) STEM images of Ta_2CSe_x. (**K** and **L**) STEM images of Ta_2CSb_x showing the similar ripple-like morphology.

Diverse surface terminations of MXenes guided by the hard and soft acid and base principle

After the redox-driven etching of A element, further oxidation of $M_{n+1}\square X_n$ by LAMS scissors (Eq. 4) results in the formation of 2D MXenes (route IV). This means that high oxidation–state M atoms in $M_{n+1}\square X_n$ could accept the nonbonding electron pair from Lewis base T^- (e.g., Cl^-, Br^- and I^-) in molten salts and form planar coordination structures (Eq. 5). Moreover, the stability of these coordination structures of MXenes is largely determined by the hard and soft acid and base (HSAB) principle when several Lewis bases coexist in a molten salt. Most of the transition metal cations (such as Ti^{4+}, Zr^{4+}, and V^{5+}) with high positive charges are typical hard Lewis acids (*26*). Consequently, the increase in chemical hardness of the halogen ligands (i.e., –I < –Br < –Cl < –F) strengthens the stability of resultant adducts, which explains the prevailing F-terminated MXenes obtained through various HF etching protocols (*10*). Although S^{2-} anion is a soft base, it could be more energetically favorable to coordinate with transition metals than Cl^-, which is consistent with the fact that the O^{2-} terminal is more stable than F^- in HF-etched MXenes. Indeed, when S^{2-} was fed by ionic compounds, FeS or CuS, into chloride melts, S-terminated MXene Ti_2CS_x and $Ti_3C_2(S_{0.5}Cl_{0.5})_x$ were obtained (fig. S26):

$$M_{n+1}\square X_n = M_{n+1}\square X_n{}^{2+} + 2e^- \quad (4)$$

$$M_{n+1}\square X_n{}^{2+} + xT^- = M_{n+1}X_nT_x \quad (5)$$

HSAB-guided coordination assists the formation of other chalcogenide MXenes (T = –Se and –Te) in molten chloride salts (figs. S27 to S31). For example, a LAMS scissor CuI etches Al out of Nb_2AlC (Eq. 2) and simultaneously oxidizes Nb atoms to a higher oxidation state (Eq. 4). The produced Cu reacts with Te ($T_m \approx 450°C$) in the chloride melt to form the ionic Cu_2Te compound with the eutectic point of 610°C (figs. S32). Last, Te^{2-} anions released from Cu_2Te and driven by electrostatic forces diffuse into positively charged $Nb_2\square C$ interlayers to form coordination with Nb atoms (Eq. 5). An accordion-like morphology is shown in the SEM image of the resultant Nb_2CTe_x MXene (Fig. 3A). The appearance of (000*l*) peaks at low Bragg angles and the disappearance of MAX-phase diffraction peaks (Fig. 3B) confirmed the complete transformation from Nb_2AlC to Nb_2CTe_x. X-ray photoelectron spectroscopy (XPS) analysis further corroborates the coordination of Te with Nb (E_{Nb3d} = 203.5 and 206.3 eV) (Fig. 3, C and D, figs. S33 and S34, and table S3) (*27*).

To form atom vacancies V_A during etching in melts, chemical scissor CuI was added in the amount sufficient for removing Al out of Nb_2AlC. A closely packed, twin-like Nb_2C structure with a typical zigzag atom arrangement was observed (Fig. 3, E and F, and fig. S35), implicating the existence of interlayer atom vacancies between Nb_2C layers, which provide the space accessible for ligand coordination and atom intercalation. Both STEM-EDS and SEM-EDS analyses semi-quantitatively identified the termination stoichiometry of $x \approx 1$ in Nb_2CTe_x, which manifests a different coordination structure with a two-electron chalcogen (Te) that bridges Nb atoms when compared with the halogen-terminated MXenes ($x \approx 2$) (fig. S36) (*13*, *19*, *21*). A ripple-like atomic arrangement appeared in Nb_2CTe_x along the *c* plane (Fig. 3, G and H, and fig. S37). This should be attributed to the lattice stress caused by large Te atoms (ionic radius = 221 pm). Actually, the lattice parameters ($a \approx 3.403$ Å and $c \approx 20.130$ Å) of Nb_2CTe_x are significantly larger than those of the parent MAX phase Nb_2AlC ($a \approx 3.106$ Å and $c \approx 13.888$ Å) (fig. S38 and table S2) (*28*). The enlarged *a* value (~9.5%

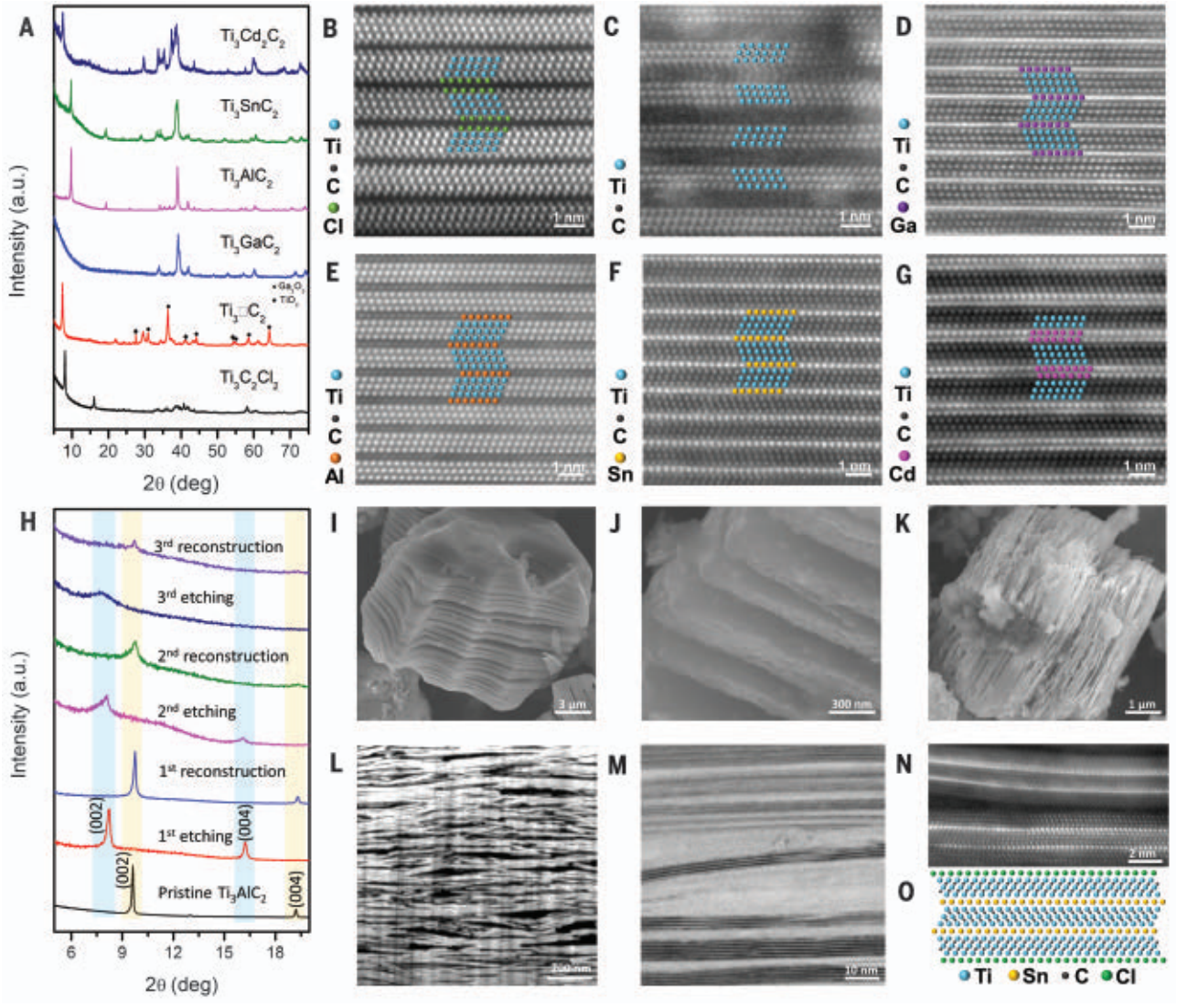


Fig. 4. Reconstruction of MAX phases from MXenes. (**A**) XRD patterns of the conversion of $Ti_3C_2Cl_2$ MXene to different MAX phases. (**B** to **G**) STEM image of $Ti_3C_2Cl_2$ MXene (B), Ga-filled $Ti_3\square C_2$ after removal of terminals (C), Ti_3GaC_2 (D), Ti_3AlC_2 (E), Ti_3SnC_2 (F), and $Ti_3Cd_2C_2$ (G) along the $[11\bar{2}0]$ zone axis. (**H**) XRD patterns showing the products after multiple etching and reconstruction. (**I** to **K**) SEM images of Ti_3AlC_2 obtained after one round (I), two rounds (J), and three rounds (K) of etching and reconstruction, respectively. (**L**) The cross-section image of the product after three rounds of etching and reconstruction, showing the stacking structure of the Ti_3AlC_2 lamellae. (**M**) Bright-field STEM image of Sn-intercalated carbide lamellae at a low magnification. (**N** and **O**) Dark-field STEM image showing atomically resolved structure of Sn-intercalated carbide $Ti_9Sn_2C_6Cl_x$ (N) and its corresponding atomic structure (O).

increase) indicates a substantial in-plane tensile stress exerted by Te on the Nb_2C layers. Indeed, the ripple-like morphology almost disappears in MXene Ta_2CSe_x (Fig. 3, I and J, and fig. S39) because Se has a smaller ionic radius of 198 pm. Furthermore, a series of mixed terminations of chalcogen and halogen, such as T = $Te_{1-x}Cl_x$, $Te_{1-x}Br_x$, and $Te_{1-x}I_x$ (x = 0 to 1), could be arbitrarily tuned by controlling the molar ratios of elements in their precursors (figs. S40 to S42).

The proposed approach can be used to expand the range of possible surface terminations. For example, phosphorus is a volatile and reactive element that cannot be directly used to functionalize MXenes. We observed that as soon as the Ti_3AlC_2 was etched by LAMS scissor $CuBr_2$, P^{3-} anions released from an ionic compound Cd_3P_2 (EP ≈ 740°C) easily attacked Ti_3C_2 together with Br^- to form a $Ti_3C_2(P_{0.4}Br_{0.6})_x$ MXene (figs. S43 and S44). Following the same exfoliation mechanism, we obtained two Sb-terminated MXenes, Ta_2CSb_x and $Ta_4C_3Sb_x$ (figs. S45 to S47). We also observed a ripple-like morphology in Ta_2CSb_x because the atomic radius of Sb is similar to that of Te (Fig. 3, K and L).

Reconstruction of 3D MAX phases from 2D MXenes enabled by metal scissors

This chemical scissor–mediated intercalation protocol is also applicable to editing MXenes. The Lewis basic ligands on MXenes can be removed by chemical scissoring of reductive metals (M′) with a low electron affinity. From the point of view of coordination, the electrons donated by reductive metals refill the unoccupied *d* orbitals of transition metal cations in MXenes (reduction reaction), reducing the effective coordination centers for the ligands, which share their electron pairs. Therefore, M atoms in $M_{n+1}\square X_n$ are reduced to a lower oxidation state, and terminations are removed from MXenes (Eq. 6) (route III). The regained nonterminated $M_{n+1}\square X_n$ provides 2D building blocks for 3D MAX phase reconstruction when guest atoms reoccupy the interlayer vacancies (Eq. 7) (route II). Taking the well-studied

$Ti_3C_2Cl_2$ MXene as an example, the metal scissor Ga removed -Cl terminations to form nonterminated $Ti_3\square C_2$ (fig. S48). The evaporation of gaseous $GaCl_3$ ($T_b \approx 201°C$) helps the complete removal of chlorine. The intercalation of guest atoms (Ga, Al, and Sn) stitched resultant $Ti_3\square C_2$ layers to reconstruct the MAX phases (Ti_3GaC_2, Ti_3AlC_2, and Ti_3SnC_2), which was confirmed by XRD patterns (Fig. 4A). Atomically resolved STEM images (Fig. 4, B to F) verify the phase conversion from $Ti_3C_2Cl_2$ MXene via nonterminated $Ti_3\square C_2$ to final reconstructed MAX phases:

$$M_{n+1}X_nT_x + M' = M_{n+1}\square X_n + M'T_x \quad (6)$$

$$M_{n+1}\square X_n + A' = M_{n+1}A'X_n \quad (7)$$

We observed wide gaps in nonterminated $Ti_3\square C_2$ when Cl terminations were removed from $Ti_3C_2Cl_2$ (Fig. 4C). EDS analysis confirmed the presence of Ga atoms distributed in the gaps, which kept the nonterminated $Ti_3\square C_2$ from collapsing into close-packed, twin-like phase (fig. S48). The gap spacing seemed large enough to accommodate more than one layer of atoms. Indeed, $Ti_3Cd_2C_2$ with a double layer of Cd (Fig. 4G) was reconstructed when metal scissor Al and intercalant metal Cd were used. Most reconstructed MAX-phase particles preserve the accordion-like morphology of multilayer MXenes (fig. S49), which indicates that the removal of terminals by metal scissors and subsequent guest atom intercalation only happens between adjacent MXene lamellas separated by no more than vdW distance.

Multiple interconversions between MAX phases and 2D MXenes could further enrich the structural editing of layered carbides. First, Ti_3AlC_2 was exfoliated by chemical scissor $CdCl_2$ to form $Ti_3C_2Cl_2$ MXene (routes I and IV). Then, the synthesized multilayer $Ti_3C_2Cl_2$ MXene was reconstructed by chemical scissor Al back into multilayer Ti_3AlC_2 (routes III and II). The characteristic (0002) diffraction peaks of $Ti_3C_2Cl_2$ and Ti_3AlC_2 confirm the successful interconversion by means of LAMS etching and metal-aided reconstruction and become substantially broadened after three cycles of interconversion because of the reduced layer thickness (Fig. 4H). The fully exfoliated Ti_3AlC_2 and $Ti_3C_2Cl_2$ lamellae were finally formed (Fig. 4, I to L, and fig. S50). The spacing between these lamellae is sufficiently large for ions to access, which may benefit diffusion-controlled electrochemical and catalytic applications. When Sn was used as an intercalant in the final reconstruction step, Ti_3SnC_2 nanosheets were built up from $Ti_3C_2Cl_2$ nanosheets (Fig. 4M). STEM images revealed that Ti_3SnC_2 nanosheets belong to a family of metal-intercalated 2D carbides in which adjacent Ti_3C_2 lamellae are intercalated by monolayers of Sn atoms but have -Cl terminations on the outmost surface (Fig. 4N and fig. S51). Therefore, such reconstructed metal-intercalated carbides combine both the functional features of MXenes that have tunable surface terminations and the structural features of MAX phases that possess oxidation-resistant interlayers. Metal-intercalated 2D carbides with the formula $M_{(n+1)m}A_{m-1}X_{nm}T_x$ (A, intercalated atom; m, number of layers) can be defined if m layers of stacked MXenes $M_{n+1}X_nT_x$ are intercalated by (m–1) layers of guest atoms after removal of (m–1) interlayer terminals by metal scissors (Eqs. 8 and 9). A Sn-intercalated 2D carbide $Ti_9Sn_2C_6Cl_x$ is obtained where $n = 2$ and $m = 3$ (Fig. 4O). If $m = 1$, then there is no intercalation at all, and the atom-intercalated 2D carbide has the same formula as MXene $M_{n+1}X_nT_x$. If m is large enough, surface terminations can be neglected for thick lamellae and the formula $M_{(n+1)m}A_{m-1}X_{nm}T_x$ is reduced to $M_{n+1}AX_n$, which represents the bottom-up reconstruction of 2D MXene nanosheets into 3D MAX phase particles:

$$mM_{n+1}X_nT_x + (m-1)M' = M_{(n+1)m}\square_{m-1}X_{nm}T_x + (m-1)M'T_x \quad (8)$$

$$M_{(n+1)m}\square_{m-1}X_{nm}T_x + (m-1)A = M_{(n+1)m}A_{m-1}X_{nm}T_x \quad (9)$$

Conclusions

The chemical scissor–mediated structural editing of MAX phases and their derived MXenes provides a powerful and versatile protocol to engineer the structure and composition of both vdW and non-vdW layered materials. The regulated intercalation routes allow the incorporation of unconventional elements into the monoatomic layer of MAX phases, which cannot be achieved through traditional metallurgic reactions, and enable the terminals' regulation of MXenes. Metal-intercalated 2D carbides, which combine the distinct structural features of MAX phases and MXenes, can be constructed through the removal of surface terminations of MXenes by metal scissors and the subsequent accommodation of guest atoms between the MXene lamellae, thus further expanding the family of layered materials. Future efforts should focus on the delamination of these 2D and 3D layered carbides, as well as metal-intercalated 2D carbides, into single- and few-layer nanosheets, which are needed for fundamental property characterization and for taking full advantage of these new materials in energy storage, electronics, and other applications.

REFERENCES AND NOTES

1. J. Zhou *et al*., *Adv. Mater.* **33**, e2004557 (2021).
2. J. Wan *et al*., *Chem. Soc. Rev.* **45**, 6742–6765 (2016).
3. M. S. Whittingham, *Science* **192**, 1126–1127 (1976).
4. T. Oshima, D. Lu, O. Ishitani, K. Maeda, *Angew. Chem. Int. Ed.* **54**, 2698–2702 (2015).
5. F. Xiong *et al*., *Nano Lett.* **15**, 6777–6784 (2015).
6. M. Burrard-Lucas *et al*., *Nat. Mater.* **12**, 15–19 (2013).
7. X. Zhao *et al*., *Nature* **581**, 171–177 (2020).
8. M. W. Barsoum, *Prog. Solid State Chem.* **28**, 201–281 (2000).
9. M. Sokol, V. Natu, S. Kota, M. W. Barsoum, *Trends Chem.* **1**, 210–223 (2019).
10. M. Naguib *et al*., *Adv. Mater.* **23**, 4248–4253 (2011).
11. M. Naguib, V. N. Mochalin, M. W. Barsoum, Y. Gogotsi, *Adv. Mater.* **26**, 992–1005 (2014).
12. A. VahidMohammadi, J. Rosen, Y. Gogotsi, *Science* **372**, eabf1581 (2021).
13. M. Li *et al*., *J. Am. Chem. Soc.* **141**, 4730–4737 (2019).
14. M. Ghidiu, M. R. Lukatskaya, M. Q. Zhao, Y. Gogotsi, M. W. Barsoum, *Nature* **516**, 78–81 (2014).
15. Y. Shao *et al*., *Nat. Commun.* **13**, 3223 (2022).
16. F. Shahzad *et al*., *Science* **353**, 1137–1140 (2016).
17. Y. Li *et al*., *ACS Nano* **13**, 9198–9205 (2019).
18. H. M. Ding *et al*., *Mater. Res. Lett.* **7**, 510–516 (2019).
19. M. Li *et al*., *ACS Nano* **15**, 1077–1085 (2021).
20. Y. Li *et al*., *Nat. Mater.* **19**, 894–899 (2020).
21. V. Kamysbayev *et al*., *Science* **369**, 979–983 (2020).
22. A. Zavabeti *et al*., *Science* **358**, 332–335 (2017).
23. X. H. Zha *et al*., *Inorg. Chem.* **61**, 2129–2140 (2022).
24. H. Fashandi *et al*., *Nat. Mater.* **16**, 814–818 (2017).
25. C. C. Lai *et al*., *Acta Mater.* **99**, 157–164 (2015).
26. T.-L. Ho, *Chem. Rev.* **75**, 1–20 (1975).
27. M. K. Bahl, *J. Phys. Chem. Solids* **36**, 485–491 (1975).
28. I. Salama, T. El-Raghy, M. W. Barsoum, *J. Alloys Compd.* **347**, 271–278 (2002).

ACKNOWLEDGMENTS

We thank P.F. Yan and X.L. Mu (Beijing University of Technology) for STEM data and the corresponding analysis. **Funding:** This work was supported by Key R&D Projects of Zhejiang Province 2022C01236 (to Q.H., K.C., and K.L.); Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang 2019R01003 (to Z.C. and Q.H.); International Partnership Program of Chinese Academy of Sciences 174433KYSB20190019 (to Q.H.); Advanced Energy Science and Technology Guangdong Laboratory HND20TDTHGC00 (to Q.H); Ningbo Top-talent Team Program and Ningbo "3315 plan" 2018A-03-A (to Z.C.); National Natural Science Foundation of China (21671195, 52172254, 52250005, 52202325, and U2004212) (to Q.H., S.D., M.L., and Y.L.); China Postdoctoral Science Foundation 2020M680082 (to Y.L.); Knut and Alice Wallenberg Foundation through the Wallenberg Academy Fellows Program KAW 2020.0196 (to P.E.) and the Wallenberg Scholar Program KAW 2019.0290 (to L.H.); the Swedish National Infrastructure in Advanced Electron Microscopy (2021-00171 and RIF21-0026) (to P.O.Å.P.); Swedish Government Strategic Research Area in Materials Science on Functional Materials at Linköping University SFO-Mat-LiU 2009 00971 (to P.O.Å.P., P.E., and L.H.); and US National Science Foundation DMR-2041050 (to Y.G.). **Author contributions:** Q.H. initiated and supervised the experimental work; Y.G. directed the discussion; H.D. conducted main experiment and analyzed results with Y.L. and M.L.; J.L., J.P., P.O.Å.P., and G.C. conducted the TEM analysis and interpretation with contributions from L.H. and P.E.; H.D. and Q.H. wrote the manuscript with contributions from all coauthors; and all authors reviewed and commented on successive drafts of the manuscript. **Competing interests:** Q.H. and H.D. are inventors on patent applications (CN202111489021.0, CN202111330269.2, CN202111488622.X, CN202111488980.0) submitted by Ningbo Institute of Materials Technology and Engineering that cover the MAX phases and MXenes described in this paper. Y.G. is also affiliated with Sumy State University, Ukraine. **Data and materials availability:** All data are available in the manuscript or the supplementary material.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add5901
Materials and Methods
Supplementary Text
Figs. S1 to S51
Tables S1 to S3
References (*29*–*46*)

Submitted 28 July 2022; accepted 16 February 2023
10.1126/science.add5901

EVOLUTION

Evolutionary transitions from camouflage to aposematism: Hidden signals play a pivotal role

Karl Loeffler-Henry[1]†, Changku Kang[2,3]*†, Thomas N. Sherratt[1]

The initial evolution of warning signals in unprofitable prey, termed aposematism, is often seen as a paradox because any new conspicuous mutant would be easier to detect than its cryptic conspecifics and not readily recognized by naïve predators as defended. One possibility is that permanent aposematism first evolved through species using hidden warning signals, which are only exposed to would-be predators on encounter. Here, we present a large-scale analysis of evolutionary transitions in amphibian antipredation coloration and demonstrate that the evolutionary transition from camouflage to aposematism is rarely direct but tends to involve an intermediary stage, namely cryptic species that facultatively reveal conspicuous coloration. Accounting for this intermediate step can resolve the paradox and thereby advance our understanding of the evolution of aposematism.

Selection to avoid being killed by predators has contributed to the diversity of animal color patterns (*1*). These color adaptations include crypsis and disruptive coloration to avoid being detected and/or recognized (*2*), conspicuous warning signals in defended species to indicate unprofitability to would-be predators [an association known as aposematism (*3*)], and mimetic signals that share or exploit the aposematic signals of other organisms (*4*). Although our understanding of the genetics, development, perception, and function of these color signals has substantially progressed in recent years (*5*), we still know little about the macroevolutionary patterns of color pattern evolution. Specifically, large-scale macroevolutionary studies on animal color defense are surprisingly scarce, and even these have tended to consider simple binary classifications of species color, notably whether they are cryptic or conspicuous (*6*–*8*). Although this binary classification captures some well-known antipredator strategies of animals (crypsis and potential aposematism or mimicry), this may not be enough to explain how diverse color defense strategies have evolved in nature. For example, some species are cryptic at rest but have bright color signals hidden on body surfaces that are only exposed when signaling to conspecifics, fleeing, or as part of a defensive posture (*9*–*14*). These flexible signaling strategies could represent intermediate stages and therefore might play a pivotal role in evolutionary processes generating different antipredator defenses (*15*).

Amphibians are an excellent group to explore the evolutionary transitions among different antipredator strategies. Their phylogeny is available in nearly all extant taxa (*16*), and their color patterns have been strongly shaped by predators through natural selection (*17*). A previous study examining macroevolutionary patterns of amphibian antipredator adaptations found that rates of speciation, extinction, and transition vary among species with different defensive traits (*6*). Although this large-scale study revealed important evolutionary pathways from camouflage to aposematism, the inference was based largely on a simple two-color state classification scheme (cryptic and conspicuous) of each species' dorsal coloration. However, many amphibian species show a more complex set of color patterns, such as having a cryptic dorsum yet conspicuous patches on normally hidden body parts (*18*–*21*). These hidden color signals are taxonomically widespread in the animal kingdom yet have seldom been considered in macroevolutionary studies (*13*, *14*).

The hidden color signals of amphibians tend to occur in one of two different forms: conspicuous color present on (i) the whole venter (lower surface), such as those in the genus *Bombina*, or (ii) part of the concealed body surface, such as ventral shanks or hindlimbs commonly found in the family Hylidae. These hidden signals are often exposed through behavioral displays [e.g., via an unken reflex toward approaching predators, or foot flagging for intraspecific signaling (*14*, *22*–*24*)]. In addition, amphibians may sometimes use flashing signals during escape: These colors would be visible only when the prey is mobile and may mislead a predator into assuming that the prey's flash color is its resting color and in so doing hinder subsequent search (*25*). Hidden conspicuous color signals may have evolved from typical aposematic signals to offset the costs of conspicuousness while stationary. Alternatively, they may serve as an intermediate state from camouflage to aposematism because this strategy can gain the advantages of both by only signaling when discovered (*15*, *26*, *27*).

Here, using discrete character evolution models, we investigate the role of hidden conspicuous color signals during the evolutionary transitions between camouflage and aposematism. Specifically, we examine the transitions between different antipredator strategies based on a five-category color classification scheme, accounting for crypsis, conspicuousness, two different types of hidden signals [PV; partially conspicuous venter: cryptic dorsum with conspicuous color present as small patches on normally hidden body parts and FV; fully conspicuous venter: cryptic dorsum with conspicuous colors that fully cover the venter (*28*)], and polymorphism (Poly; defined here as a species having both cryptic and conspicuous forms regardless of whether they are regional variants or coexist in the same population; see Fig. 1 for example species and materials and methods for details). We identified two classes of hidden signals because (i) the classes are distinguishable morphologically and (ii) their putative functions differ in that FV coloration is likely to solely function as an aposematic signal to attacking or approaching predators, whereas PV coloration may also serve as a flashing signal or territorial display (*14*). We analyzed two datasets: color (1106 species with color information available) and color+chem (315 species with both color and chemical defense information available).

Results and discussion

Main evolutionary transitions in the color model

We tested nine different models: three models that allowed transitions between almost all states (with two exceptions, see below) at equal or differing rates [All rates different (ARD), Symmetric (SYM), Equal rates (ER) models], and six models that restricted certain transitional pathways (Intermediate, FV intermediate, Stepwise, PV/FV secondary, FV cost-offset PV secondary, Cost-offset) (table S1 and fig. 1A). Transitions between the FV and Poly states and between the PV and Poly states were not allowed (see materials and methods for details). An intermediate model that did not allow the direct transition of species between the cryptic and conspicuous states was the best-supported (lowest Akaike information criterion) model (Fig. 1A and tables S1 and S2 for descriptions and full results), while the estimated transition rates were qualitatively equivalent among the three best-supported models [the Intermediate, ARD, and a modified Intermediate model in which the PV state cannot evolve directly to the conspicuous state but only through the FV state (FV intermediate); fig. S1]. The parameter estimates from fitting the Intermediate model revealed several major patterns of evolutionary pathways

[1]Department of Biology, Carleton University, Ottawa, Ontario K1S 5B6, Canada. [2]Department of Agricultural Biotechnology, Seoul National University, Seoul 08826, South Korea. [3]Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 08826, South Korea.
*Corresponding author. Email: changkukang@snu.ac.kr
†These authors contributed equally to this work.

(Fig. 1B). First, although species in the cryptic state can evolve directly toward all other states with the exception of the conspicuous state (which is precluded in the best-supported model), the cryptic state is stable (i.e., the transition rates from all other states to the cryptic state are at least three times higher than the rates away from it) and the most likely basal ancestral state (support probability = 69%; Fig. 2). Second, the PV state is most strongly associated with the cryptic state: Species with the PV color mainly evolve from cryptic coloration and tend to transit back to it. However, although the transition rate is low, the PV state is also the most likely state leading to the FV state, which could facilitate the subsequent transition toward the conspicuous state. Third, species in the conspicuous state largely evolve from the FV or polymorphic state, and other routes are less likely. However, because species in the polymorphic state evolve almost exclusively from the conspicuous state, the major pathway to the initial evolution of conspicuous coloration is likely through the FV state. Fourth, the polymorphic state appears unstable and transits rapidly to the cryptic or conspicuous state.

Main evolutionary transitions in the color+chem model

In the color+chem model, the evolutionary pathways among the states are substantially more complex than the color-only models (Fig. 1C). As with the color model, the cryptic state was found to be the most likely basal ancestral color state in the subset of species considered in the color+chem model (support probability = 60%; Fig. 3). Although there was slightly stronger support for the hypothesis that the basal defensive state of frogs and salamanders involved chemical defense, the data were more equivocal (Fig. 3; 58% support probability for chemical defense versus 42% for no chemical defense). Chemical defense frequently coincides with conspicuous coloration: Indeed, the majority of species classified as either conspicuous, FV color, or polymorphic have chemical defense (more than 90% in all three groups; table S3). However, chemical defense also occurs in less-conspicuous species as well (65% in species with PV color, 51% in the cryptic species; table S3). This is not surprising, as chemical defense is known to provide survival advantages to both conspicuous and cryptic species through aposematic signaling and taste rejection, respectively (*29*–*31*). On average, the transition rates toward the acquisition of defense are higher than the transition rates away from this defensive state, but the only color state that is more likely to lose than acquire chemical defense is crypsis (Fig. 1C). This is consistent with previous findings that the acquisition of alkaloid sequestration is favored over losing it in poison frogs (*32*). One explanation for the loss of chemical defense in cryptic amphibians is that they experience less risk of detection by predators and therefore less selective pressure for the maintenance of postdetection defense (*33*). Considering that chemical defense may pose a cost to its bearer (*34*, *35*), this may make deterrent toxins less favored in some cryptic lineages. Another nonmutually exclusive reason for the differential in acquisition or loss of chemical defenses in the different types of signaler may be that species with any form of conspicuous signals can potentially benefit from "go slow" signaling that makes predators cautiously sample the prey to determine the presence of chemical defense (*36*). By contrast, predators may sample cryptic prey without caution, possibly making selection for chemical defense to reduce injury after capture weaker in cryptic species.

After accounting for chemical defense, the evolutionary transition from crypsis to aposematism (chemically defended conspicuous state) is not simple but is instead composed

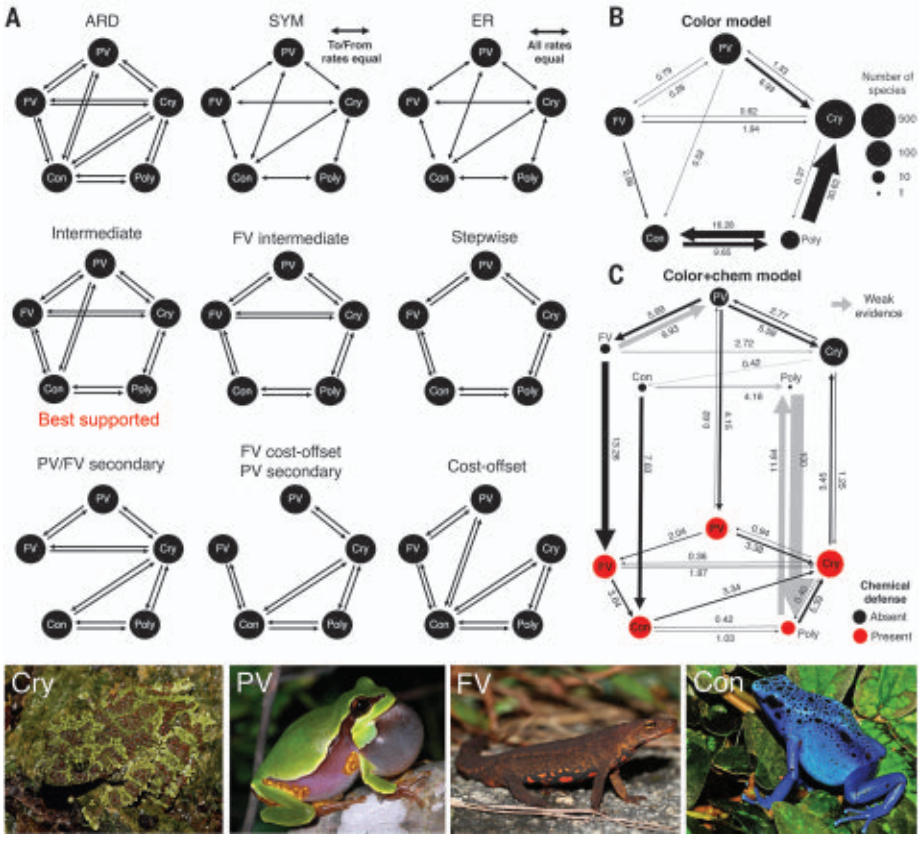


Fig. 1. Diagrammatical representations of the models that we compared and the parameter estimates of the best-supported models. No arrows in the fitted transition models (**A**) indicate that the transition rates were constrained to be zero. See table S1 for the descriptions of the models. Panels (**B**) and (**C**) show the estimated transition rates among different states from the best-supported models using color dataset [(B); species with color information available; *N* = 1106] and color+chem dataset [(C); species with both color and chemical defense information available; *N* = 315 species]. Species that were classified as uncertain color (*N* = 47) were excluded. Arrow thickness reflects the estimated transition rates, which are also given by the values listed next to each arrow (those transitions that are possible but not depicted have estimated rates less than 0.00001). The radius of circles is proportional to the log-transformed number of species in each state. Gray arrows indicate that the strength of evidence is weak because the estimated transition rates were inconsistent between different functions ("fitMk" and "corHMM") (*28*), most likely because they were estimated from very few changes in the tree or they were estimated from only one extant species (undefended Poly). Cry: cryptic; PV (partially conspicuous venter): cryptic dorsum with conspicuous color present as small patches on normally hidden body parts; FV (fully conspicuous venter): cryptic dorsum with conspicuous colors fully covered on the venter; Con: conspicuous; Poly: a species showing multiple distinct morphs that include both cryptic and conspicuous forms. ARD: All rates different model; SYM: symmetric model; ER: equal rates model (see table S1 for details). The photos show sample species from each color category that we used. From left to right, *Theloderma corticale* (Dan Rosenberg), *Hyla andersonii* (Troy Hibbitts), *Paramesotriton hongkongensis* (Dan Rosenberg), *Dendrobates tinctorius* (Michael Gäbler).

of multiple pathways that involve intermediate states (Fig. 1C; see fig. S2 for the visual comparisons of how accounting for chemical defenses modifies the transition rates among the color states from the color model results). Whereas the transition from the undefended conspicuous state to the defended conspicuous state is high, the transition rates toward the undefended conspicuous state from any other states are either zero or very low. This is reasonable because any mutations that make an individual conspicuous without having chemical defense would be detrimental and as such would be selected against unless other defensive strategies, such as Batesian mimicry, are involved (*37*, *38*). Thus, a more stable route to aposematism is via the chemically defended FV state. There are multiple pathways to evolve the chemically defended FV state, but the main routes appear through the defended PV state or undefended FV state, which also mainly evolve from the undefended PV state. These observations collectively suggest that at least the FV state (but often involving the PV state) is likely to be involved in transitions from camouflage to aposematism. Once aposematism has evolved, it goes back to neither the PV nor FV state but instead either evolves back to the cryptic state directly or becomes cryptic/conspicuous-mixed polymorphic.

Hidden signals and their implications for amphibian color evolution

We hypothesized that the PV signals potentially function as a secondary defense (flash displays or postdetection warning) or are used for intraspecific signaling in normally cryptic species (*14*, *25*); thus, the PV signals are mainly associated with cryptic coloration. Our results of both color and color+chem models support this view in that the PV state is primarily associated with the cryptic state with a stronger tendency to go back to the cryptic state (Fig. 1, B and C). Also, the PV state is the most likely state that can lead to the evolution of the FV state, which is a major precursory state toward conspicuous coloration. The presence of PV signals implies that a species has managed to express bright (e.g., carotenoid or pteridine pigments based) colors potentially acquired through diet and/or manufactured de novo and presumably evolved for antipredator or conspecific signaling (*14*, *39*, *40*). This could further facilitate the expression of conspicuous colors on other parts of the body, resulting in the evolution of the FV states.

In both color and color+chem models, species having FV color evolve from either the cryptic or PV color, but not from the conspicuous color (Fig. 1, B and C). Thus, the conjecture that the FV color evolves from the conspicuous color to offset the cost of being continuously conspicuous is not supported (*15*). Instead, the FV state is the most likely intermediate stage that is required for the transition from crypsis to aposematism. About 91% of species with FV color have chemical defense (table S3), suggesting that their ventral warning coloration is likely an honest signal of their defense, rather than a bluff. Theoretically, the FV state can have a selective advantage over the conspicuous state when a species has no chemical defense: Having a conspicuous dorsum without defense should be highly detrimental to individuals, leaving less opportunity to evolve chemical defense subsequently. However, because the FV strategy does not involve the loss of crypsis, this strategy may be able to persist until the evolution of chemical defense follows. Indeed, the results of the color+chem model suggest that the nonchemically defended conspicuous state rarely evolves from any other nonchemically defended states, but the FV state could evolve from the PV state in the absence of chemical defense (Fig. 1C).

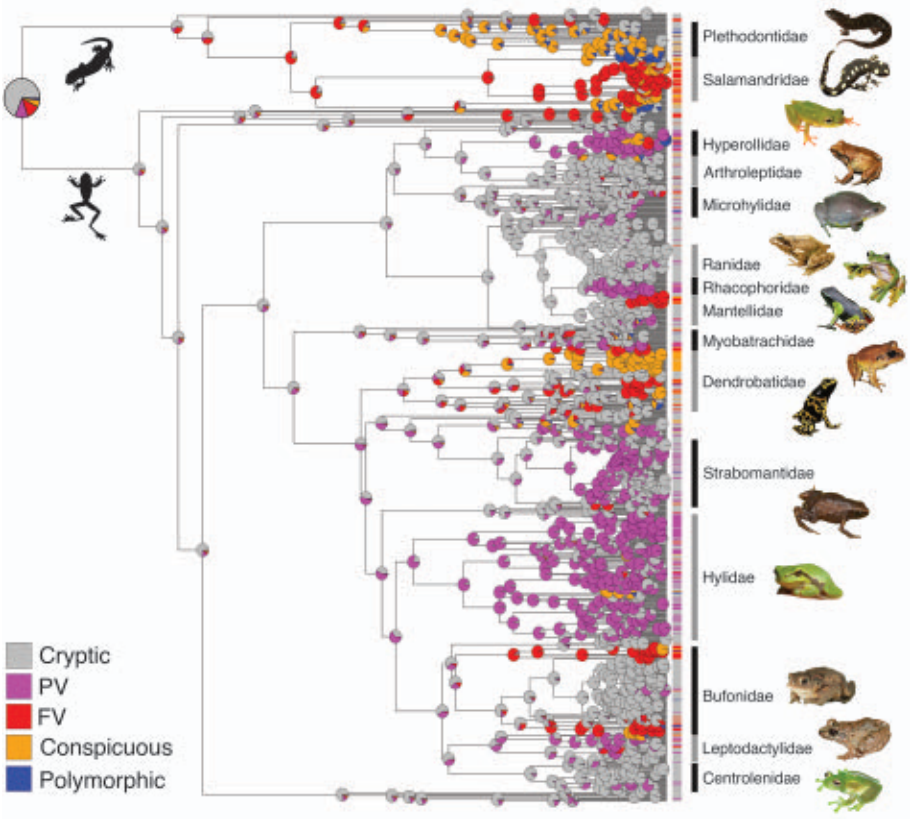


Fig. 2. Ancestral state estimation of each color state (*N* = 1106 species) in frogs and salamanders. Pie charts at each node show the probabilities of ancestral states. The ancestral state of frogs and salamanders is likely to be cryptic coloration. The hidden color signals (PV and FV) are widespread and have evolved multiple times in different lineages. PV: cryptic dorsum with conspicuous color present as small patches on normally hidden body parts; FV: cryptic dorsum with conspicuous colors fully covered on the venter. See table S11 for photo credits.

Transitional patterns of the cryptic/conspicuous-mixed polymorphic state

Only 2.3% of all species in our dataset were considered to exhibit both cryptic and conspicuous morphs, i.e., were classified as polymorphic. Despite its relative rarity, our results suggest that this cryptic/conspicuous-mixed polymorphism has evolved multiple times in different lineages independently (Fig. 2). Most of these polymorphic species have chemical defense (10 out of 11 species) and have evolved mainly from the conspicuous states (Fig. 1, B and C). Both the color and color+chem models suggest that once the cryptic/conspicuous-mixed polymorphic state is reached, it tends to rapidly evolve toward either the cryptic or conspicuous state with a stronger tendency toward the cryptic state (Fig. 1, B and C). The low number of species in the mixed-polymorphic state may reflect this evolutionary instability.

Conclusion

Our study highlights the importance of hidden color signals for the evolutionary processes that generate diverse antipredator coloration in amphibians. Our results suggest that (i) species with hidden color signals, especially those with conspicuous colors that cover the whole venter, represent a key stage in the evolution of aposematic species from cryptic

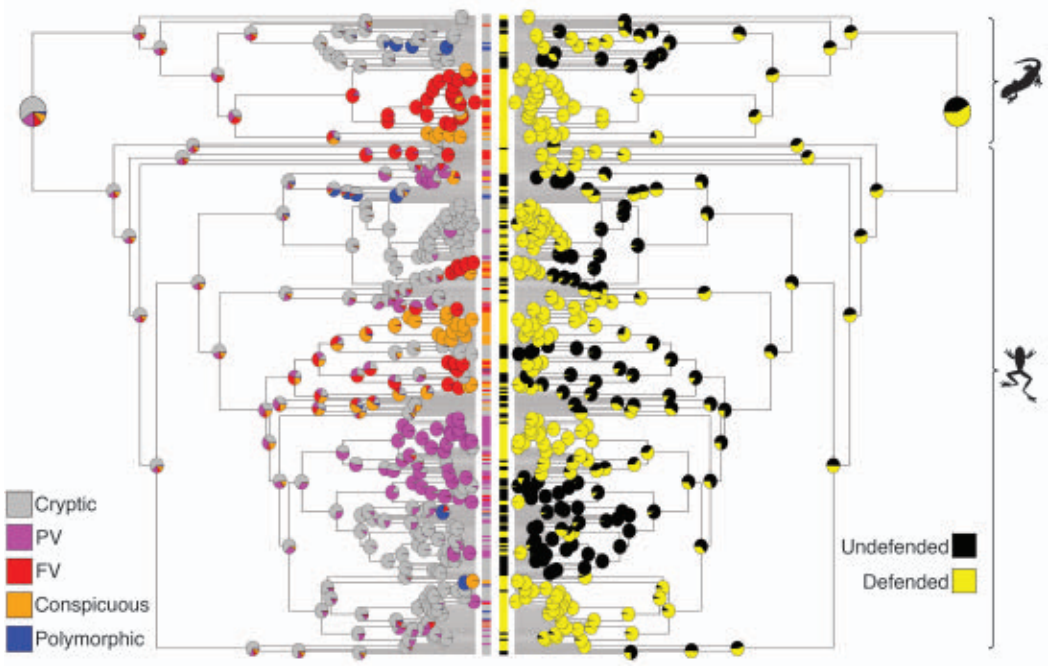


Fig. 3. Ancestral state estimation of each combination of color and chemical defense in frogs and salamanders (*N* = 315 species). Pie charts at each node show the probabilities of ancestral states. The ancestral state of frogs and salamanders is likely to be cryptic coloration, but the evidence of whether the basal state was chemically defended or not was equivocal. Most states have evolved multiple times in different lineages. Transitions from cryptic to aposematic (chemically defended conspicuous) states have usually occurred via intermediate states.

species; (ii) cryptic/conspicuous-mixed polymorphism plays a pivotal role in transitions from aposematism back to crypsis; and (iii) the transition rates toward acquisition of chemical defense are higher than those to its loss in most color states, with the exception of the cryptic state. A number of complementary models confirmed the robustness of these conclusions (see materials and methods for details and figs. S3 to S9 and tables S4 to S10 for the results).

Biologists have long wondered how rare conspicuous mutants of a cryptic defended species can spread in a population when they have a higher predation risk before predators learn to avoid them (*41*, *42*). The fact that aposematic species appear to seldom derive directly from cryptic species confirms the long-standing intuition that the evolution of aposematism from crypsis is challenging. Here we show macroevolutionary evidence for an important yet unrecognized solution in which the problem is effectively side-stepped: Rather than evolve from cryptic species, aposematic mutants appear to derive largely from species with hidden signals. One intuitive way that this might work is if would-be predators that are already exposed to a species with hidden warning signals continue to treat permanently aposematic mutants of this species with caution. The initial evolution of hidden color signals might be facilitated by predatory pressure that promotes the evolution of secondary defenses such as flash or warning displays (*15*, *24*, *25*), or via sexual selection that results in the expression of conspicuous signals in body parts that are only visible during behavioral displays (*14*). These complex transitional routes from crypsis to aposematism would not be revealed if traditional dichotomous classifications of animal antipredator coloration (either cryptic or conspicuous) are applied (see fig. S10 for the results when the binary classification was used). Thus, macroevolutionary studies on animal coloration should take into account these underappreciated hidden signals, which are both common and widespread across the animal kingdom (*13*, *43*, *44*), to advance our understanding of the evolution of antipredator defenses. Indeed, many animal taxa such as snakes, fishes, and a variety of arthropods (see fig. S12 for example groups) include species that are cryptic, are aposematic, and have hidden conspicuous signals. We therefore encourage follow-up studies in other taxa to evaluate the generality of the stepping-stone hypothesis as a route to aposematism.

REFERENCES AND NOTES

1. H. B. Cott, *Adaptive Coloration in Animals* (Methuen, 1940).
2. M. Stevens, S. Merilaita, *Animal Camouflage: Mechanisms and Function* (Cambridge Univ. Press, 2011).
3. E. B. Poulton, *The Colours of Animals: Their Meaning and Use, Especially Considered in the Case of Insects* (D. Appleton, 1890).
4. G. D. Ruxton, A. L. William, T. N. Sherratt, M. P. Speed, *Avoiding Attack: The Evolutionary Ecology of Crypsis, Warning Signals, and Mimicry* (Oxford Univ. Press, ed. 2, 2018).
5. I. C. Cuthill *et al.*, *Science* **357**, eaan0221 (2017).
6. K. Arbuckle, M. P. Speed, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13597–13602 (2015).
7. B. B. Blaimer, J. R. Mawdsley, S. G. Brady, *Evolution* **72**, 1874–1889 (2018).
8. K. Przeczek, C. Mueller, S. M. Vamosi, *Integr. Zool.* **3**, 149–156 (2008).
9. M. Vidal-García, J. C. O'Hanlon, G. J. Svenson, K. D. L. Umbers, *Proc. Biol. Sci.* **287**, 20201016 (2020).
10. M. Edmunds, *Zool. J. Linn. Soc.* **51**, 1–32 (1972).
11. T. Caro, T. N. Sherratt, M. Stevens, *Evol. Ecol.* **30**, 797–809 (2016).
12. W. Hödl, A. Amézquita, in *Anuran Communication*, M. J. Ryan, Ed. (Smithsonian Institution Press, 2001), pp. 121–141.
13. K. Loeffler-Henry, C. Kang, T. N. Sherratt, *Am. Nat.* **194**, 28–37 (2019).
14. I. Starnberger, D. Preininger, W. Hödl, *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **200**, 777–787 (2014).
15. K. D. L. Umbers *et al.*, *Biol. Lett.* **13**, 20160936 (2017).
16. W. Jetz, R. A. Pyron, *Nat. Ecol. Evol.* **2**, 850–858 (2018).
17. B. Rojas, *Biol. Rev. Camb. Philos. Soc.* **92**, 1059–1080 (2017).
18. J. Bajger, *Herpetologica* **36**, 133–137 (1980).
19. A. Rudh, A. Qvarnström, *Semin. Cell Dev. Biol.* **24**, 553–561 (2013).
20. M. Martins, *J. Herpetol.* **23**, 305–307 (1989).
21. L. F. Toledo, C. F. B. Haddad, *Int. J. Zool.* **2009**, 910892 (2009).
22. L. F. Toledo, I. Sazima, C. F. B. Haddad, *Ethol. Ecol. Evol.* **23**, 1–25 (2011).
23. D. P. Ferraro, M. O. Pereyra, P. E. Topa, J. Faivovich, *Zool. J. Linn. Soc.* **193**, 388–412 (2020).
24. R. B. Ferreira *et al.*, *Behav. Ecol. Sociobiol.* **73**, 69 (2019).
25. K. Loeffler-Henry, C. Kang, Y. Yip, T. Caro, T. N. Sherratt, *Behav. Ecol.* **29**, 528–533 (2018).
26. T. Guilford, *Am. Nat.* **131**, S7–S21 (1988).
27. M. P. Speed, G. D. Ruxton, *Proc. Biol. Sci.* **272**, 431–438(2005).
28. Materials and methods are available as supplementary materials.
29. J. Mappes, N. Marples, J. A. Endler, *Trends Ecol. Evol.* **20**, 598–603 (2005).
30. Y. Yoshimura, E. Kasuya, *PLOS ONE* **8**, e81280 (2013).
31. J. Skelhorn, C. Rowe, *Behav. Ecol. Sociobiol.* **60**, 550–555 (2006).
32. J. D. Carvajal-Castro, F. Vargas-Salinas, S. Casas-Cardona, B. Rojas, J. C. Santos, *Sci. Rep.* **11**, 19047 (2021).
33. L. Wang, G. D. Ruxton, S. J. Cornell, M. P. Speed, M. Broom, *J. Theor. Biol.* **473**, 9–19 (2019).
34. M. D. Bowers, in *Insect Chemical Ecology: An Evolutionary Approach*, B. D. Roitberg, M. B. Isman, Eds. (Chapman and Hall New York, 1992), pp. 216–244.
35. R. A. Blennerhassett, K. Bell-Anderson, R. Shine, G. P. Brown, *Proc. Biol. Sci.* **286**, 20190867 (2019).
36. T. Guilford, *J. Theor. Biol.* **170**, 311–316 (1994).
37. C. R. Darst, M. E. Cummings, *Nature* **440**, 208–211 (2006).
38. L. Lindström, R. V. Alatalo, A. Lyytinen, J. Mappes, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9181–9184 (2001).
39. K. D. L. Umbers, A. J. Silla, J. A. Bailey, A. K. Shaw, P. G. Byrne, *Biol. J. Linn. Soc. Lond.* **119**, 436–444 (2016).
40. T. Goodwin, *The Biochemistry of the Carotenoids, Volume I, Plants* (Springer Science & Business Media, 2012).
41. R. A. Fisher, *The Genetical Theory of Natural Selection: A Complete Variorum Edition* (Oxford Univ. Press, 1930).
42. R. V. Alatalo, J. Mappes, *Nature* **382**, 708–710 (1996).

43. T. Caro, H. Raees, T. Stankowich, *Behav. Ecol. Sociobiol.* **74**, 44 (2020).
44. K. Loeffler-Henry, C. Kang, T. N. Sherratt, *Proc. Biol. Sci.* **288**, 20210866 (2021).
45. K. Loeffler-Henry, C. Kang, T. N. Sherratt, Amphibian color paper data and analysis code. *figshare*. Dataset. https://doi.org/10.6084/m9.figshare.19890313 (2023).

ACKNOWLEDGMENTS

We are grateful to three anonymous reviewers who provided a plethora of insightful feedback during the peer review process for this manuscript. We thank E. Kerr and J. Butler from the Canadian Herpetological Society for assistance in classifying each species.

Funding: C.K. is supported by the National Research Foundation of Korea (grant no. NRF-2019R1C1C1002466) and the New Faculty Startup From Seoul National University. T.N.S. and K.L-H. are supported by the Natural Sciences and Engineering Research Council of Canada (NSERC, Discovery Grant to T.N.S.). **Author contributions:** C.K. conceived the study. K.L-H. conducted the image collecting and classifications. C.K. conducted the phylogenetic analysis with input from K.L-H. and T.N.S. All authors contributed to discussions and interpretations of the data. All authors contributed to formulating the initial draft of the manuscript and subsequent revisioning. **Competing interests:** The authors declare no competing interests. **Data and material availability:** All data and analysis codes have been deposited at Figshare (*45*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade5156
Materials and Methods
Figs. S1 to S12
Tables S1 to S12
References (*46*–*70*)

Submitted 23 August 2022; accepted 21 February 2023
10.1126/science.ade5156

NEURODEGENERATION

Mechanism of *STMN2* cryptic splice-polyadenylation and its correction for TDP-43 proteinopathies

Michael W. Baughn[1,2]†, Ze'ev Melamed[1,2,3]*†, Jone López-Erauskin[1,2], Melinda S. Beccari[1,2], Karen Ling[4], Aamir Zuberi[5], Maximilliano Presa[5], Elena Gonzalo-Gil[5], Roy Maimon[1,2], Sonia Vazquez-Sanchez[1,2], Som Chaturvedi[2], Mariana Bravo-Hernández[2], Vanessa Taupin[2], Stephen Moore[2], Jonathan W. Artates[1,2], Eitan Acks[1,2], I. Sandra Ndayambaje[6], Ana R. Agra de Almeida Quadros[6], Paayman Jafar-nejad[4], Frank Rigo[4], C. Frank Bennett[4], Cathleen Lutz[5], Clotilde Lagier-Tourenne[6,7]*, Don W. Cleveland[1,2]*

Loss of nuclear TDP-43 is a hallmark of neurodegeneration in TDP-43 proteinopathies, including amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). TDP-43 mislocalization results in cryptic splicing and polyadenylation of pre–messenger RNAs (pre-mRNAs) encoding stathmin-2 (also known as SCG10), a protein that is required for axonal regeneration. We found that TDP-43 binding to a GU-rich region sterically blocked recognition of the cryptic 3′ splice site in *STMN2* pre-mRNA. Targeting dCasRx or antisense oligonucleotides (ASOs) suppressed cryptic splicing, which restored axonal regeneration and stathmin-2–dependent lysosome trafficking in TDP-43–deficient human motor neurons. In mice that were gene-edited to contain human *STMN2* cryptic splice-polyadenylation sequences, ASO injection into cerebral spinal fluid successfully corrected *Stmn2* pre-mRNA misprocessing and restored stathmin-2 expression levels independently of TDP-43 binding.

In the human nervous system, the ability to maintain proper RNA metabolism is thought to decline during aging (*1*). Disruption of RNA metabolism is a common feature of many human neurodegenerative disorders, including the fatal paralytic disease amyotrophic lateral sclerosis (ALS) and the two most common dementias: Alzheimer's disease (AD) and frontotemporal dementia (FTD) (*1*–*5*). Although the exact gene expression profiles and affected neuronal populations vary among the dementias and other neurodegenerative disorders, there is growing evidence supporting common molecular mechanisms (*6*, *7*).

TDP-43 proteinopathy describes a set of neurological disorders that are characterized by mislocalization of the RNA-binding protein TDP-43 [encoded by the *TARDBP* (TAR DNA-binding protein) gene]. TDP-43 relocates from its typically nuclear location and accumulates in the cytoplasm of affected neurons in the form of aggregates. TDP-43 mislocalization and aggregation is found in 97% of ALS patients, about half of FTD patients, and 30 to 50% of AD patients (*8*–*11*). TDP-43 pathology has also been reported in a growing number of brain disorders (*2*), including Huntington's disease (*12*), Perry syndrome (*13*), and chronic traumatic encephalopathy (CTE) (*14*). Recently, a subset of aged Alzheimer's patients has been reclassified as limbic-predominant age-related TDP-43 encephalopathy (LATE disease) (*15*) because their post mortem brain samples show aberrant TDP-43 instead of the expected amyloid beta (Aβ). Although cytoplasmic accumulation of TDP-43 has been reported in ALS and FTD, nuclear clearance of TDP-43 is often observed even without apparent aggregation(s) (*16*, *17*). Loss of nuclear TDP-43 affects expression and processing of multiple mRNA targets across different cell types and tissues (*18*–*24*).

TDP-43 closely regulates maturation of the pre-mRNA encoding stathmin-2 (also known as SCG10, encoded by the *STMN2* gene) (*25*, *26*). Stathmin-2 is a neuronally enriched protein that plays a crucial role in axonal outgrowth during development (*27*) and regeneration (*25*, *26*, *28*). Developmental deletion of mouse *Stmn2* causes motor deficits with denervation of neuromuscular junctions (*29*, *30*). Among the four members of the stathmin gene family, stathmin-2 has the highest expression in mouse and human motor neurons (*25*); *STMN2* is among the top 20 most enriched mRNAs in the ALS-vulnerable motor neurons of the anterior gray column of the spinal cord (*25*, *31*). In post mortem FTD and ALS patient brain and spinal cord tissues, loss of nuclear TDP-43 results in the use of cryptic splice (*25*, *26*, *32*) and polyadenylation sites (*25*) within the first intron of the *STMN2* pre-mRNA (*25*, *26*). This leads to inclusion of a cryptic exon 2a and the production of an mRNA encoding only a truncated open reading frame (17 codons) (*25*) and suppression of the full-length 179–amino acid stathmin-2 protein (*25*, *26*, *32*).

Stathmin-2 is a tubulin-binding protein that is thought to affect microtubule dynamic instability (*33*, *34*), although its mechanism of action in axons and axonal growth cones is unclear (*35*–*38*). Lowering stathmin-2 in induced pluripotent stem cell (iPSC)–derived human motor neurons inhibits the regenerative capacity of injured axons (*25*, *26*). Diminished regeneration capacity of injured motor neurons when TDP-43 function is reduced can be rescued by restoring stathmin-2 expression levels (*25*), which suggests that restoration of stathmin-2 expression in TDP-43 proteinopathies may provide a therapeutic strategy.

In this work, we determined the regulatory elements through which TDP-43 regulates *STMN2* pre-mRNA processing and identified steric binding antisense oligonucleotides (ASOs)

[1]Ludwig Institute for Cancer Research, University of California at San Diego, La Jolla, CA 92093, USA. [2]Department of Cellular and Molecular Medicine, University of California at San Diego, La Jolla, CA 92093, USA. [3]Department of Medical Neurobiology, Faculty of Medicine, The Hebrew University of Jerusalem, Israel. [4]Ionis Pharmaceuticals, Carlsbad, CA 92010, USA. [5]Rare Disease Translational Center, The Jackson Laboratory, Bar Harbor, ME 04609, USA. [6]Department of Neurology, Sean M. Healey & AMG Center for ALS, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA. [7]Broad Institute of Harvard University and MIT, Cambridge, MA 02142, USA.
***Corresponding author. Email: dcleveland@health.ucsd.edu (D.W.C.); clagier-tourenne@mgh.harvard.edu (C.L.-T.); zmelamed@mail.huji.ac.il (Z.M.)**
†These authors contributed equally to this work.

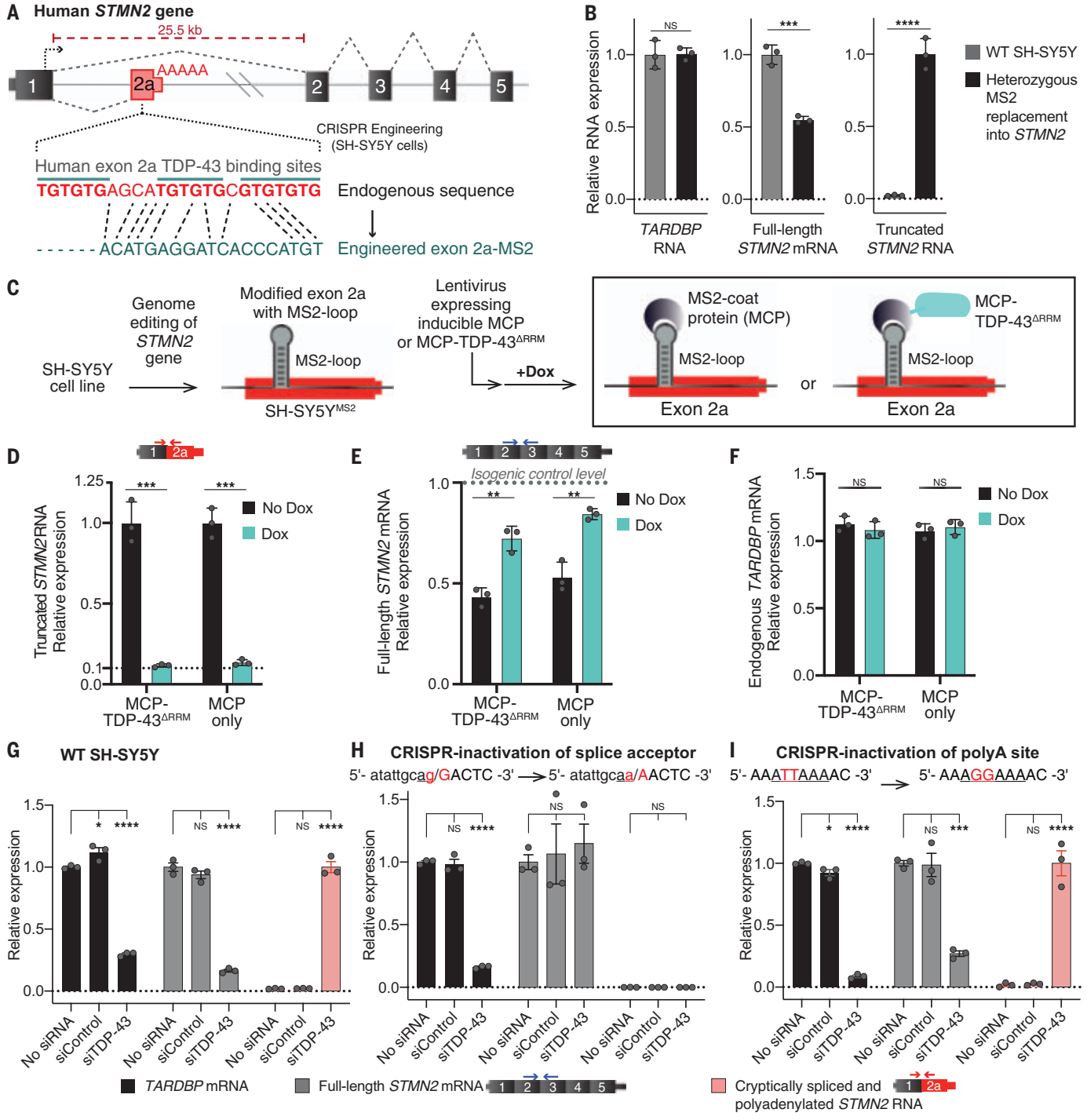


Fig. 1. Human GU-motif removal and MS2-directed tethering demonstrate TDP-43 binding locus, and cryptic site mutations identify TDP-43–dependent misprocessing requiring cryptic splice acceptor. (**A**) Schematic of CRISPR-engineering strategy for conversion of the GU binding motif in exon 2a into an MS2 aptamer sequence in one *STMN2* allele of diploid SH-SY5Y neuroblastoma cells. (**B**) Quantitative reverse transcription polymerase chain reaction (RT-PCR) demonstrating that SH-SY5Y cells carrying heterozygous GU to MS2 edit misprocess *STMN2* RNA, leading to 50% loss of stathmin-2–encoding mRNA compared with that in wild-type cells and accumulation of truncated RNA. (**C**) Schematic depicting MS2:MCP–directed strategy to direct MCP-tethered proteins to the normal TDP-43 binding locus. (**D** to **F**) Quantitative RT-PCR measurement of (D) truncated *STMN2*, (E) full-length *STMN2* mRNA, or (F) endogenous *TARDBP* mRNA levels with and without induction of MCP-fusion protein expression in SH-SY5Y cells carrying heterozygous MS2 aptamer insertion. (**G** to **I**) Quantitative RT-PCR measured expression of *TARDBP*, full-length *STMN2* mRNA, and truncated *STMN2* RNAs 96 hours after siRNA treatment with a control siRNA pool or a pool targeting *TARDBP* in (G) wild-type SH-SY5Y cells, (H) SH-SY5Y cells harboring homozygous mutation of the human exon 2a 3′ splice acceptor site, and (I) SH-SY5Y cells harboring a homozygous mutation of the human exon 2a premature polyadenylation signal to the mouse sequence AGGAAA. For all quantitative PCR analysis, individual data points are independently treated wells of cells. Error bars are SEM. Statistical significance was determined by means of two-tailed Student's *t* test in (B) to (F), or one-way analysis of variance (ANOVA) with Dunnett correction in (G) to (I). ****P <0.0001; ***P < 0.001; **P < 0.01; *P < 0.05.

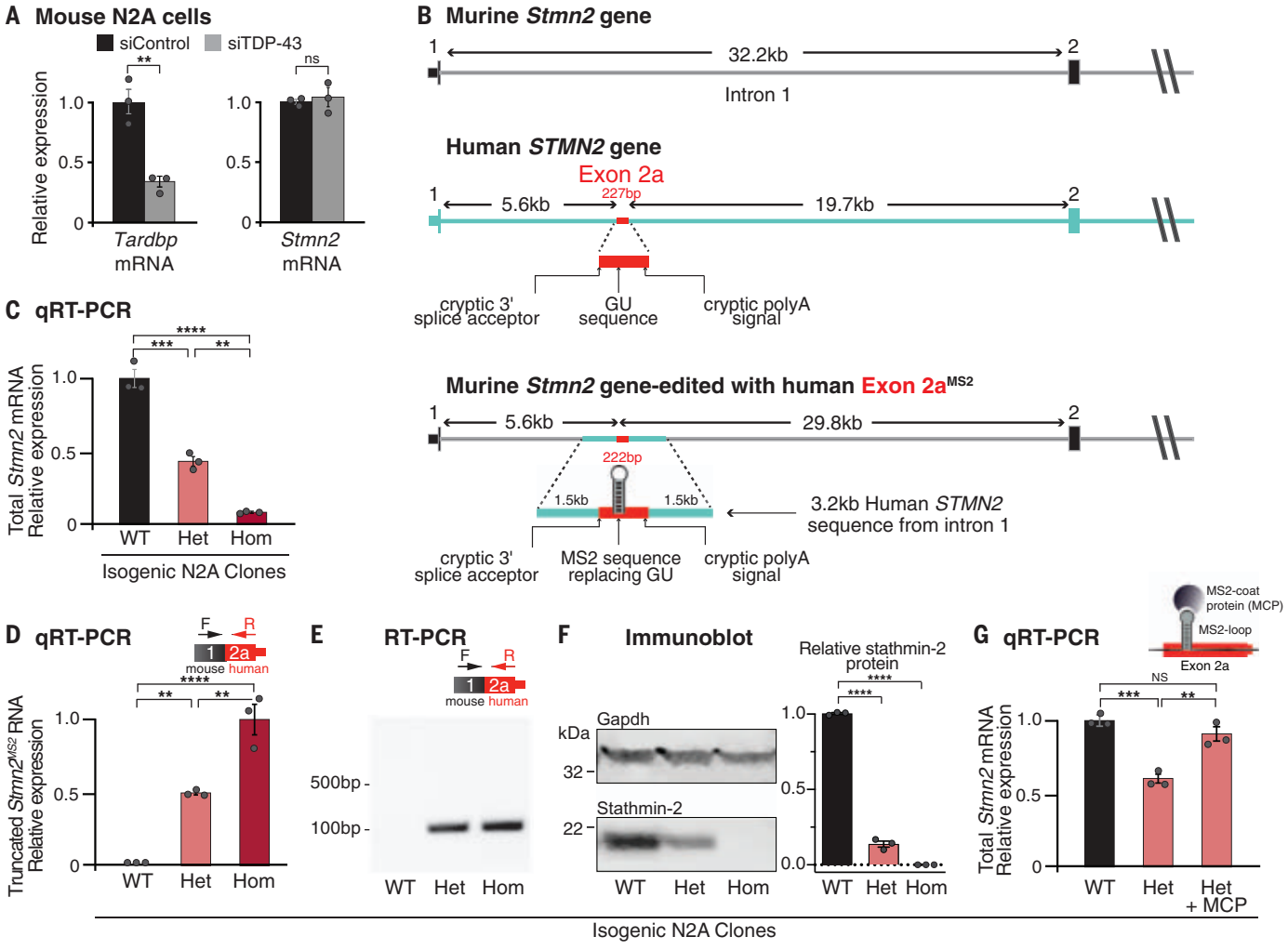


Fig. 2. Humanization of mouse Stmn2 gene in N2A cells demonstrates human-specific inhibition of altered STMN2 pre-mRNA processing via non-conserved TDP-43 binding sites. (**A**) Quantitative RT-PCR demonstrating that depletion of TDP-43 in wild-type mouse N2A cells does not affect *Stmn2* expression levels. (**B**) Schematics of the human and mouse *Stmn2* genomic regions before and after genome editing to insert a 3-kb human fragment of *STMN2* intron 1 into mouse N2A cells. (**C**) Quantitative RT-PCR demonstrating dose-dependent reduction of N2A *Stmn2* mRNA level that correlates with the number of alleles carrying human *STMN2* gene fragment. (**D**) Quantitative RT-PCR and (**E**) RT-PCR confirming expression of chimeric *Stmn2* truncated RNA with mouse exon 1 spliced to human exon 2a, in N2A clones carrying the human *STMN2* gene fragment. (**F**) Immunoblotting confirming reduced expression of full-length stathmin-2 protein levels in N2A clones that carry humanized *Stmn2* gene fragment. (**G**) Quantitative RT-PCR showing restoration of normal *Stmn2* pre-mRNA processing in N2A cells upon doxycycline induction of MCP expression. For all quantitative PCR analyses, each data point indicates an independently treated well of N2A cells. Error bars are SEM. Statistical significance was determined by means of two-tailed Student's *t* test in (A) or one-way ANOVA with Dunnett correction in (F) or Tukey correction in (C), (D), and (G). $^{****}P < 0.0001$; $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$.

capable of restoring normal stathmin-2 protein and RNA levels when administered within a mammalian nervous system.

A GU-motif confers TDP-43 binding–dependent maturation of STMN2 pre-mRNA

A previous analysis of datasets for RNAs bound by TDP-43 [using ultraviolet individual nucleotide resolution crosslinking, immunoprecipitation, and sequencing (iCLIP)] identified TDP-43 binding within the human (*18*) but not mouse (*19*) *STMN2* pre-mRNA, ~5.6 kb from the start (5′ end) of the first *STMN2* intron (*25*). The iCLIP-defined TDP-43 binding site is positioned between the cryptic splice and polyadenylation sites of the recently identified exon 2a (*25*), in a region containing a 24-base GU-rich segment comprising three closely spaced GUGUGU hexamers, which are the consensus motif for TDP-43 binding (*19*, *39*). We used CRISPR-Cas9 genome engineering to test whether TDP-43 binding at this locus could prevent misprocessing by blocking recognition of the cryptic RNA elements. We replaced the 24-base domain that encodes the GU-motif sequence with a 19-base segment that encodes the bacteriophage MS2 aptamer sequence, an RNA stem-loop structure that can be bound with high affinity [dissociation constant (K_d) = 2×10^{-10} M] (*40*) by the MS2 coat protein (MCP) (Fig. 1, A to C). In human neuronal SH-SY5Y cells carrying the MS2-binding site replacement in one *STMN2* allele (fig. S1A), steady-state full-length *STMN2* mRNAs were reduced by 50%. This was accompanied by appearance of an abundant, truncated *STMN2* RNA produced by use of the cryptic splice and polyadenylation sites, despite sustained TDP-43 levels (Fig. 1B). Thus, the GU motif bound by TDP-43 was required to suppress use of cryptic sites within the *STMN2* pre-mRNA.

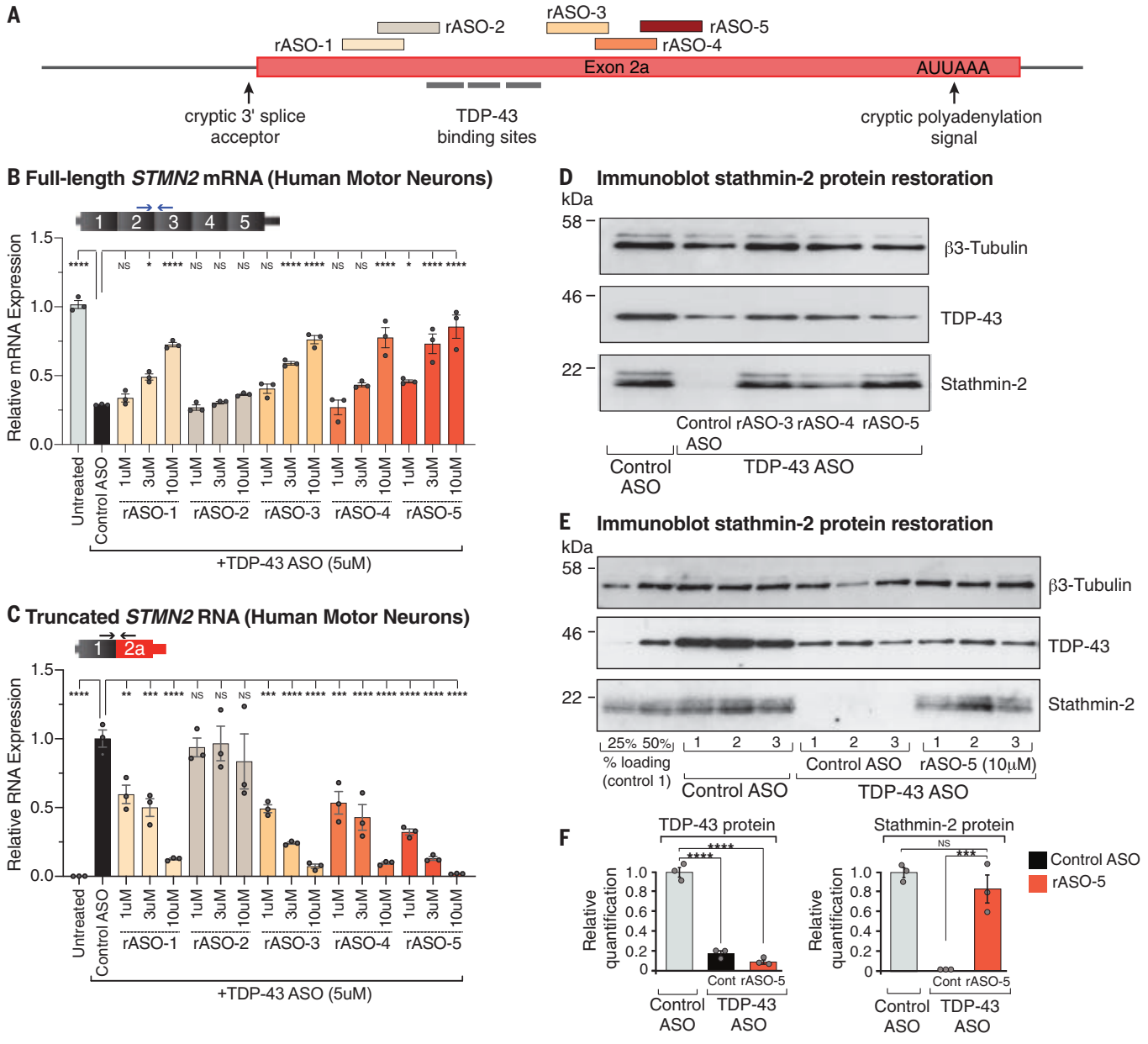


Fig. 3. Dose-dependent suppression of STMN2 cryptic splicing and polyadenylation by rASOs in iPSC-derived motor neurons with TDP-43 depletion. (**A**) Schematic representation of the exon 2a region of human *STMN2* gene with TDP-43 binding sites and selected rASOs that show splice-modifying activity. (**B**) *STMN2* mRNA restoration analyzed by means of quantitative RT-PCR after treatment with five representative rASOs in iPSC-derived motor neurons depleted of TDP-43. Expression of *TFRC* mRNA was used as endogenous control. (**C**) Quantitative RT-PCR analysis of truncated *STMN2* RNA levels after treatment with five representative rASOs in iPSC-derived motor neurons depleted of TDP-43. Expression of *TFRC* mRNA was used as endogenous control. (**D**) Immunoblot showing TDP-43 and stathmin-2 protein levels in iPS motor neurons treated with control or TDP-43–suppressing ASOs, subsequently treated with control or splice-rescuing rASOs 3, 4, or 5 to restore stathmin-2. β3 Tubulin was used as an endogenous control. (**E**) Immunoblot showing TDP-43 and stathmin-2 levels in motor neurons depleted of TDP-43 with a targeted ASO and subsequently treated with control or rescue ASO 5. Linearity of antibody detection with 25 and 50% control-treated motor neuron lysate loading controls are included at the far left side of the blot. Immunoblots are quantified in (**F**). Each lane and data point indicates an independently differentiated and ASO-treated neuronal culture. Error bars are SEM. Statistical significance was determined by means of one-way ANOVA with Dunnett correction. ****P < 0.0001; ***P < 0.001; **P < 0.01; *P < 0.05.

TDP-43 binding sterically blocks cryptic splicing and polyadenylation of STMN2

TDP-43 binding to the GU motifs could act directly to prevent use of cryptic processing sites in the *STMN2* pre-mRNA by sterically blocking accessibility to those sites through additional splicing (*41*–*44*) or polyadenylation (*45*–*48*) factors. Alternatively, TDP-43 binding to the GU motifs could act indirectly by promoting recruitment and assembly of additional RNA processing protein complexes whose presence occludes access of 3′ splice site or transcription termination and polyadenylation complexes. To distinguish between direct and indirect models for suppressing cryptic site use in the *STMN2* pre-mRNA, we transduced our SH-SY5Y cells, in which the MS2 aptamer was inserted into one *STMN2* allele, with a lentivirus carrying

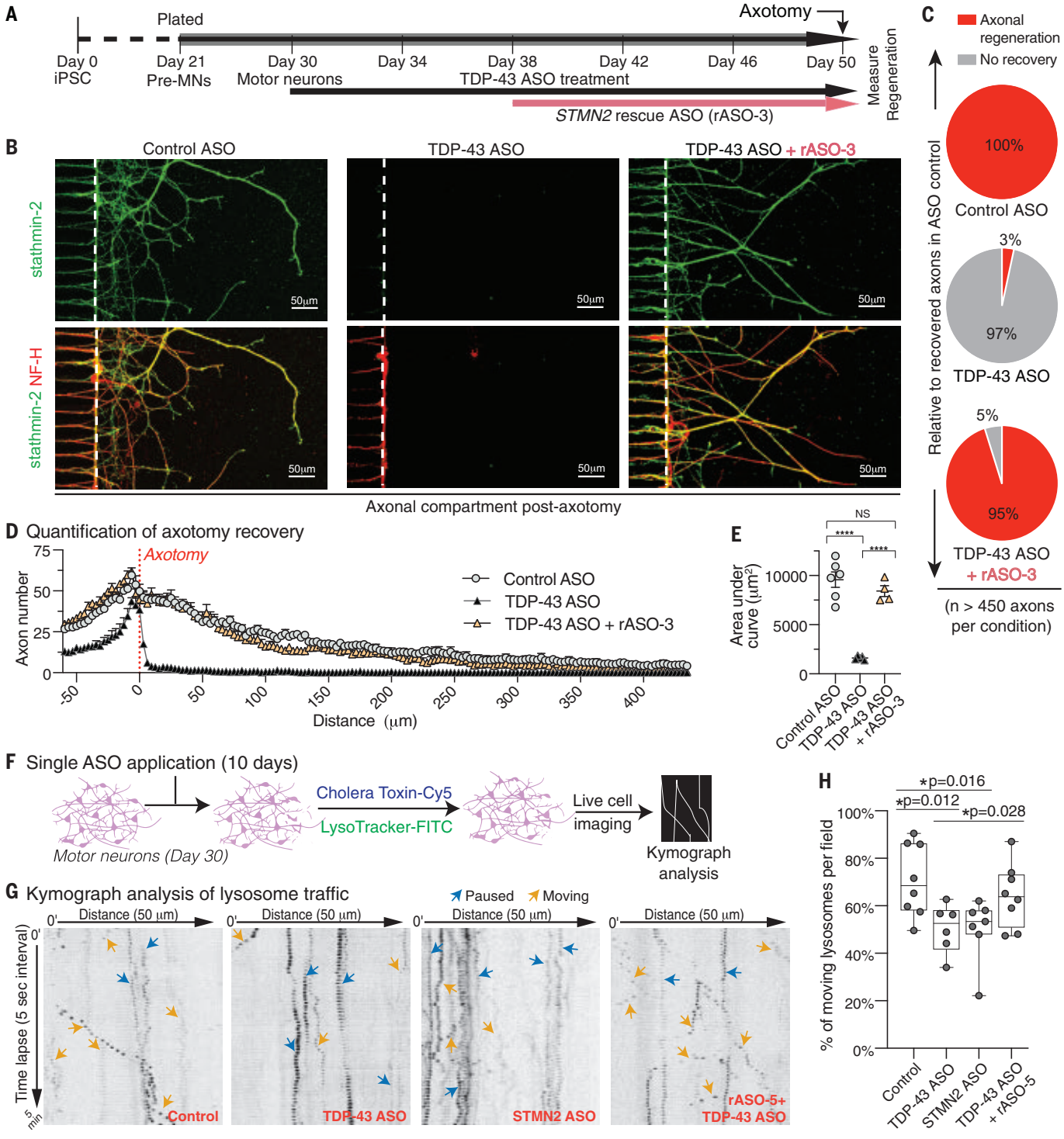


Fig. 4. Restoration of axonal regeneration capacity by using rASOs that rescue stathmin-2 levels in iPSC-derived motor neurons with TDP-43 depletion. (**A**) Timeline of iPSC-derived motor neuron maturation, ASO or rASO treatment, and axotomy. (**B**) Immunofluorescence images of microgrooves (left of dotted line) and distal compartments (right), 36 hours after axotomy. Axonal regeneration and growth cones are observed by means of immunofluorescence detection of stathmin-2 (green) and NF-H (red) in the terminals of motor neurons. (**C** to **E**) Quantification of axonal recovery for at least 450 axons per condition represented as (C) the percentage of recovered axons relative to control ASO-treated motor neurons, (D) the overall number of axons per micrometer from the axotomy site plotted, and (E) corresponding area-under the curve. Statistical significance was determined by means of one-way ANOVA with Tukey's multiple comparison correction. (**F**) Schematic of ASO treatment and live motor neuron lysosomal tracking and analysis. (**G**) Representative kymographs of lysosomal transport in axons of ASO-treated motor neurons. (**H**) Quantification of moving tracked axonal lysosomes after ASO treatment. Statistical significance was determined by means of one-way ANOVA indicated, with three independently differentiated chambers quantified per condition. Error bars are SEM. ****$P < 0.0001$; ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$.

a doxycycline-inducible gene encoding a 29-kDa high-affinity MS2 coat protein dimer fused to an RNA-binding-incompetent human TDP-43 variant (MCP-TDP-43$^{\Delta RRM}$) missing both RNA-recognition motifs (ΔRRM). This allowed us to direct single-molecule binding of the fusion protein to the *STMN2*-embedded MS2 RNA aptamer but not to TDP-43 binding sites on other endogenous pre-mRNAs (Fig. 1C). Within 24 hours of MCP-TDP-43$^{\Delta RRM}$ induction, cryptically spliced and polyadenylated stathmin-2 mRNAs were reduced by 90% (Fig. 1D, left). Next, to determine the effect of simple steric binding at the replaced GU locus, we induced expression of MCP alone (without TDP-43 fusion) (Fig. 1C). Targeting exon 2a by means of MCP binding at the MS2 sequence similarly reduced accumulation of truncated *STMN2* RNA by 90% (Fig. 1D, right). Correspondingly, full-length stathmin-2 mRNA levels were largely restored through binding of either MCP variant (Fig. 1E), whereas endogenous *TARDBP* mRNA levels remained unaltered (Fig. 1F). Thus, TDP-43 binding to the GU-domain of the *STMN2* pre-mRNA blocked cryptic splicing and polyadenylation through a simple steric inhibition mechanism.

Genomic disruption of exon 2a cryptic splice site protects against misprocessing

We next tested whether recognition of either the cryptic splice or polyadenylation sites (which lie 203 bases from each other within the *STMN2* pre-mRNA) were responsible for initiation of *STMN2* pre-mRNA misprocessing when TDP-43 levels fall. We used genome editing in SH-SY5Y cells to eliminate either the cryptic polyadenylation motif or the 3′ splice acceptor site (fig. S1, B and C). As previously reported (*25*), small interfering RNA (siRNA)–mediated reduction of TDP-43 to 30% of its initial level in the parental cells resulted in an even larger (85%) reduction in stathmin-2–encoding mRNAs (Fig. 1G). In cells homozygously CRISPR-edited to convert the cryptic 3′ splice acceptor sequence from a functional AG/GA to a nonfunctional AA/AA sequence, siRNA reduction of TDP-43 resulted in no significant change in full-length *STMN2* mRNA and complete abrogation of both cryptic splicing and polyadenylation (Fig. 1H).

By contrast, elimination of the identified cryptic polyadenylation signal was not protective against *STMN2* pre-mRNA misprocessing. In cells homozygously edited to carry a disrupted cryptic polyadenylation signal [after converting the initial ATTAAA sequence to a nonfunctional AGGAAA (*49*)] (Fig. 1I), reduction in TDP-43 sharply reduced full-length *STMN2* RNAs (Fig. 1I, gray bars) and generated RNAs with exon 1 ligated onto exon 2a (Fig. 1I, pink bars). Moreover, despite inactivation of the most proximal site of exon 2a polyadenylation, cryptically spliced *STMN2* RNAs remained polyadenylated. A search of putative cryptic polyadenylation sites downstream within the intron 1 sequence identified 39 occurrences of the canonical AATAAA signal for polyadenylation and 22 additional instances of the ATTAAA variant, one or more of which must become new cryptic polyadenylation sites when the first ATTAAA was inactivated.

Cryptic human STMN2 splice and pA sites induce misprocessing of mouse *Stmn2*

Despite 93% conservation in mRNA coding sequence and 100% conservation at the amino acid level, the mouse *Stmn2* gene has little conservation of sequence within the region of its first intron corresponding to human exon 2a, including absence of predicted or experimentally validated TDP-43 binding site(s) (*19*). Correspondingly, cryptic *Stmn2* splicing and polyadenylation were not present in the mouse neuron-like N2A cell line after TDP-43 level was reduced by siRNA and full-length *Stmn2* mRNA level remained unchanged (Fig. 2A).

We genome engineered one or both *Stmn2* alleles in these mouse cells by insertion of a 3.2-kb segment containing the exon 2a sequence of human *STMN2* intron 1 after a 12-nucleotide modification to convert the GU-rich TDP-43 binding motif into an MS2 aptamer (Fig. 2B). The resultant edited mouse *Stmn2* alleles contained human exon 2a sequence positioned as in the human gene, 5.6 kb downstream of the 5′ splice site (5′ss) of intron 1. Heterozygous or homozygous insertions of the cryptic splicing and polyadenylation sites from human *STMN2* resulted, respectively, in reduction or nearly complete elimination of mature mouse *Stmn2* mRNA (Fig. 2C), driven by constitutive use of the human cryptic 3′ splice and polyadenylation sites, leading to production of a chimeric mouse-human truncated *Stmn2* RNA (Fig. 2, D and E). Mouse stathmin-2 protein was accordingly reduced or eliminated in heterozygous and homozygous clones, respectively (Fig. 2F). Lentiviral-mediated expression of MCP restored levels of full-length mouse *Stmn2* mRNA (Fig. 2G). Thus, the cryptic human *STMN2* 3′ splice acceptor and polyadenylation elements could drive misprocessing of mouse *Stmn2* pre-mRNAs when TDP-43 was not bound.

Targeting of dCasRx blocks misprocessing of STMN2 pre-mRNA

The CRISPR effector RfxCas13d (CasRx) can be targeted to a specific RNA sequence with an appropriate ≥21-base guide RNA (*50*). The "nuclease-dead" variant of which (dCasRx) retains RNA-binding but not enzymatic cleavage activity and can be directed to pre-mRNA molecules to affect alternative splicing (*50*). We applied an initial test in human neuronal cells of the therapeutic potential in TDP-43 proteinopathies of dCasRx-directed targeting to sterically block use of cryptic splice and polyadenylation sites within the *STMN2* pre-mRNA. This possibility is especially attractive for dCasRx given its small size and reported versatility in RNA binding that is independent of a protospacer adjacent motif (PAM) targeting requirement (*50*). Eleven guide RNAs were designed with sequences that tiled across a 353-base region spanning the cryptic splice and polyadenylation sites in the human *STMN2* pre-mRNA (fig. S2A). SH-SY5Y cells were genome edited to carry an ALS- or FTD-linked TDP-43$^{N352S/N352S}$ mutation within both endogenous *TARDBP* alleles (*25*), resulting in a partial loss of TDP-43 function, production of a truncated *STMN2* RNA (fig. S2, B and C), and a corresponding 50% reduction of the full-length stathmin-2–encoding mRNAs (Fig. S2B,D).

These cells were then transduced with a lentivirus encoding dCasRx and each of the 11 gRNAs from our tiling array. Four guides with binding sites covering (i) the cryptic 3′ splice acceptor site (gRNA 3), (ii) overlapping or (iii) directly adjacent to the endogenous TDP-43 binding site (gRNA 6 and 7), and (iv) covering the cryptic polyadenylation signal (gRNA 10)–generated (fig. S2A) marked reduction of up to 70% in the levels of cryptically spliced, polyadenylated *STMN2* RNA relative to a control TDP43$^{N352S/N352S}$ line expressing dCasRx without a gRNA (fig. S2, B and C). gRNA-mediated targeting of dCasRx to partly overlap the authentic TDP-43 binding site (gRNA 6) showed an up to 40% increase in levels of full-length stathmin-2–encoding mRNAs, as did direct targeting of cryptic splice acceptor or polyadenylation signals (fig. S2, B and D). Guides directing binding to an intermediate locus upstream to the alternative polyadenylation signal produced further depletion of normal *STMN2* levels in these cells (fig. S2B, gRNA 8 and 9), perhaps displacing an additional RNA-binding factor with a role in pre-mRNA maturation. Stathmin-2 protein expression was restored proportionally to the mRNA levels, with gRNA 6 mediating 30% protein increase relative to no guide control (fig. S2, E and F). Thus, suppression of cryptic splicing and polyadenylation and partial restoration of *STMN2* expression in cells with diminished TDP-43 activity was achieved through dCasRx-mediated steric binding on the exon 2a sequence.

Antisense oligonucleotide suppression of STMN2 pre-mRNA misprocessing

Antisense oligonucleotides (ASOs) that have been chemically modified to correct RNA processing defects without recruiting ribonuclease H (RNase H) to catalyze RNA degradation have become an attractive therapeutic approach for neurodegenerative diseases (*51*, *52*). Their usage to alter pre-mRNA splicing of the survival of motor neuron 2 (*SMN2*) gene (*51*) is an approved standard of care for spinal muscular atrophy (SMA) in the United States and

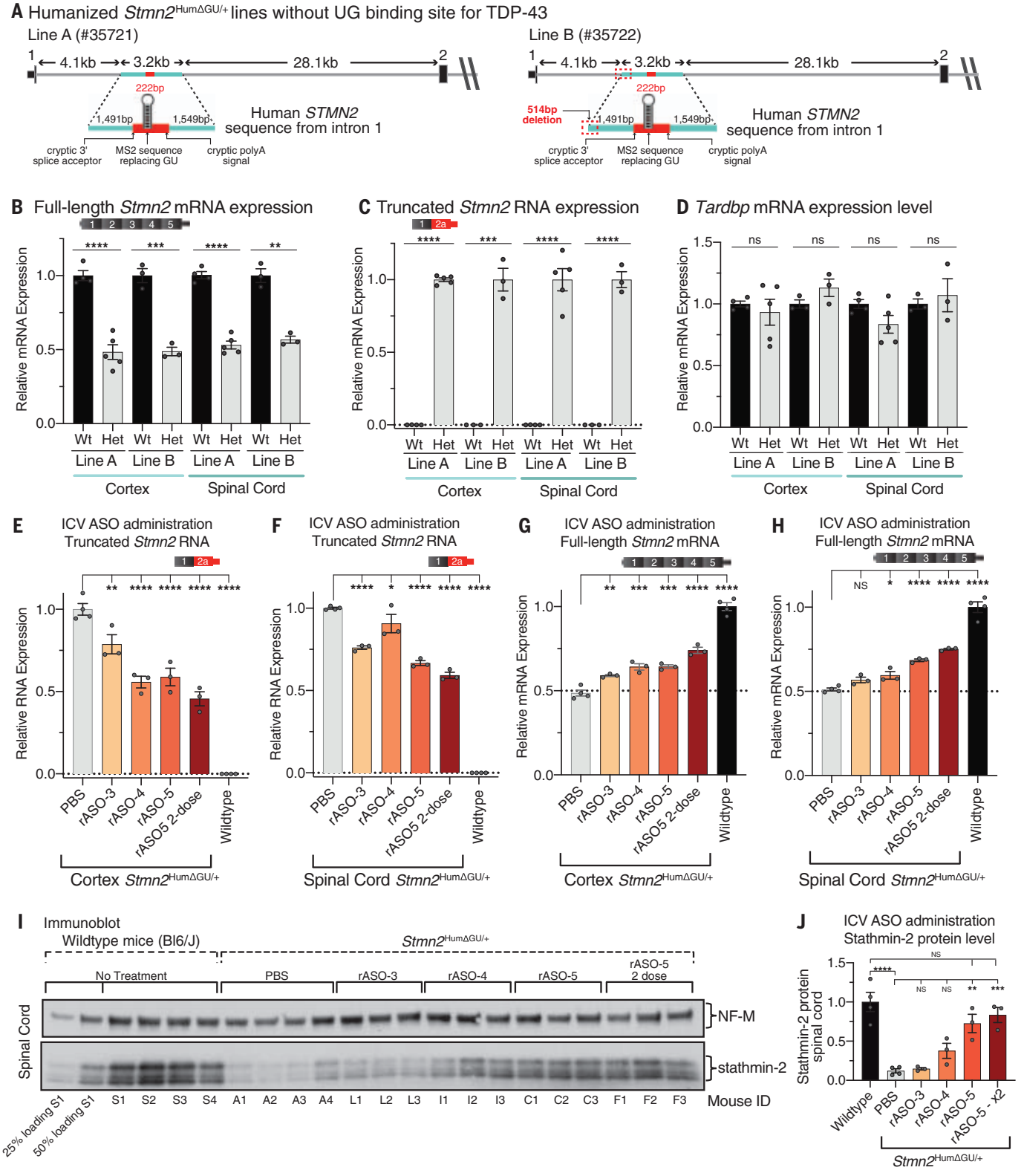


Fig. 5. In vivo ASO-mediated restoration of *Stmn2* pre-mRNA processing in a humanized mouse, engineered to constitutively misprocess *Stmn2* RNAs from a partly humanized allele. (**A**) Schematic showing the strategy to produce a mouse in which *Stmn2* human exon 2A was inserted without a TDP-43 GU binding site to drive constitutive misprocessing of the humanized allele, and resulting mouse lines obtained after CRISPR editing. (**B**) Quantitative RT-PCR showing full-length mouse *Stmn2* mRNA levels were reduced by half in animals heterozygous for the humanized $Stmn2^{HumΔGU/+}$ allele compared with wild-type littermate controls. (**C**) Quantitative RT-PCR showing that truncated chimeric RNAs consisting of a mouse exon 1 fused to a modified human exon 2a were abundantly and specifically expressed in mice heterozygously carrying the modified humanized $Stmn2^{HumΔGU/+}$ allele. (**D**) Quantitative RT-PCR showing normal

mouse *Tardbp* mRNA levels in both heterozygous humanized *Stmn2*$^{HumΔGU/+}$ mice and wild-type littermate controls. (**E** and **F**) Quantitative RT-PCR showing suppression of truncated chimeric RNA accumulation in both (E) brain and (F) spinal cord of mice dosed by means of ICV injection with rASOs. (**G** and **H**) Quantitative RT-PCR showing rASO-mediated restoration of full-length mouse *Stmn2* mRNAs in both (G) brain and (H) spinal cord of mice after ICV injection of rASOs. (**I**) Immunoblot showing restoration of stathmin-2 protein in the spinal cords of heterozygous humanized *Stmn2*$^{HumΔGU/+}$ mice dosed by means of ICV injection with rASOs, compared with phosphate-buffered saline (PBS)–injected animals or wild-type C57 BL/6J mice. Twenty-five and 50% loading of lysates from wild-type mouse S1 are at the far left, and NF-M is shown as an endogenous loading control protein. (**J**) Quantification of relative stathmin-2 protein restoration. Each lane and data point indicates an individual mouse. Error bars are SEM. Statistical significance was determined by means of two-tailed Student's *t* test in (B) to (D), or one-way ANOVA with Dunnett correction in (E) to (H) or Tukey correction in (J). ****$P < 0.0001$; ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$.

Europe. Correspondingly, 250 potential rescue ASOs (rASOs) were synthesized to target across the region spanning cryptic splice and polyadenylation sites in the *STMN2* pre-mRNA (fig. S3A) and evaluated for their ability to suppress cryptic splicing and polyadenylation of *STMN2* pre-mRNAs. Using SH-SY5Y cells harboring a homozygous TDP-43$^{N352S/N352S}$ mutation, we identified several rASOs that within 24 hours produced dose-dependent restoration of full-length stathmin-2–encoding mRNAs to a level comparable with that in wild-type cells (fig. S3B) and an accompanying reduction in levels of the truncated stathmin-2 RNA variant (fig. S3C). Correspondingly, rASOs-mediated blockage of cryptic splicing and polyadenylation resulted in stathmin-2 protein restoration as early as 48 hours after transfection (fig. S3, D and E).

rASOs restore axonal regeneration capacity in TDP-43–deficient motor neurons

To determine the therapeutic potential of rASOs in the neuronal population most affected in ALS, we generated motor neurons from induced pluripotent stem cells, reduced TDP-43 levels through addition of an ASO targeting *TARDBP* RNAs (*25*), and assayed for restoration of *STMN2* mRNA and protein levels after addition to cell culture media of *stathmin-2* rASOs (Fig. 3A). Reduction in TDP-43 resulted in accumulation of truncated *STMN2* RNAs and suppression of full-length *STMN2* mRNAs to 30% of their initial level (Fig. 3, B and C). In the absence of a rASO, stathmin-2 protein levels dropped to an undetectable level in TDP-43–depleted neurons (Fig. 3, D to F). rASO treatment increased *STMN2* mRNAs in a dose-dependent manner in TDP-43–depleted motor neurons, with rASO-5 elevating *STMN2* mRNA from 30% to an average of 86% of its normal level (Fig. 3B), whereas the truncated *STMN2* RNA accumulation was almost completely reversed (Fig. 3C). Encouragingly, stathmin-2 protein level, which was below the detection limit in motor neurons with ASO-depletion of TDP-43, was rescued by rASOs (Fig. 3D), with rASO-5 returning stathmin-2 protein to a level nearly indistinguishable from that in control neurons with normal TDP-43 levels (Fig. 3, E and F).

To determine the functional consequences of rASO-mediated stathmin-2 restoration on the regenerative capacity of motor axons, we used an iPSC-motor neuron axotomy and regrowth assay in microfluidic chambers (Fig. 4A) (*25*). Addition into the somatic compartment of an ASO targeting TDP-43–encoding mRNAs for RNase H–dependent degradation for 20 days led to loss of detectable stathmin-2 protein (Fig. 4B, middle, green) accompanied by inhibition of axonal regeneration capacity after mechanically induced axotomy (Fig. 4, B and C, middle), in contrast to motor neurons treated with nontargeting control ASOs (Fig. 4, B and C, left and top, respectively). Relevance of stathmin-2 in axonal regeneration is established through direct reduction of stathmin-2 (by using a *STMN2*-targeting ASO); loss of stathmin-2 alone is sufficient to almost eliminate axonal regeneration after axotomy (*25*).

Free uptake of an rASO added to the somatic compartment 8 days after initiation of ASO-dependent reduction in TDP-43–restored accumulation of axonal stathmin-2 protein levels (Fig. 4B, right, green). The ability of axotomized motor axons to regenerate into the distal compartment was also rescued, with nearly identical recovered axon number and density into the distal compartment when compared with those of control axons (Fig. 4B, right; quantified in Fig. 4, C to E), despite apparent heterogeneity in stathmin-2 accumulation in individual axons (Fig. 4B). Analysis of regenerating axons after ASO-mediated reduction in stathmin-2 revealed that a residual level of stathmin-2 protein, 25% of the normal level, was sufficient to support axonal regeneration after injury in human motor neurons (fig. S4, A to C).

Recognizing that impaired axonal transport has been demonstrated in several ALS models (*53*–*56*), we next applied live-cell imaging to track organelle movement within cultured iPSC-derived motor neurons (Fig. 4F). Movement of lysosomes in iPSC-derived motor neurons treated with ASOs to degrade *STMN2* or *TARDBP* mRNAs showed a decrease in actively moving lysosomes (reduction from 71% in axons exposed to control ASOs to 50% after stathmin-2 or TDP-43 suppression) (Fig. 4, G and H). rASO treatment to restore stathmin-2 levels significantly reversed the impaired lysosome trafficking phenotype despite sustained TDP-43 suppression, as evident by an increased proportion (to 64%; $P = 0.028$) of lysosomes that were moving (Fig. 4H).

We then used electron microcopy imaging to provide ultrastructural analysis of excitatory synapses under conditions of stathmin-2 or TDP-43 suppression. Overall, synaptic morphology did not show abnormalities (fig. S5, A to E), including active zone and post-synaptic density (PSD) length and total amount of synaptic vesicles. However, in TDP-43 and/or stathmin-2–deficient motor neurons, multiple synapses showed an increase of electron dense material in the active zone, in comparison with that of motor neurons treated with a control ASO (fig. S5, A to F). rASO treatment to specifically restore stathmin-2 levels in motor neurons, despite continuous TDP-43 suppression, fully restored the proportion of synapses showing normal presynaptic electron density to reflect that of control neurons (fig. S5F, top and bottom bars).

rASO-mediated restoration of *Stmn2* pre-mRNA processing in the mammalian CNS

Recognizing that mouse *Stmn2* pre-mRNA processing is independent of TDP-43 function (Fig. 2A) owing to absence of the regulatory sequences that define human exon 2a, we tested whether humanizing the mouse *Stmn2* pre-mRNA by introduction of human exon 2a and flanking intronic sequences was sufficient to confer TDP-43 dependency. For this, we generated humanized *Stmn2* mice from mouse embryonic stem cells that had been CRISPR-Cas9 genome-edited to replace a 479-bp segment of mouse *Stmn2* intron 1 with a 394-bp segment of human *STMN2* intron 1 containing exon 2a (227 bp) and its flanking regions (75 bp upstream and 92 bp downstream). Exon 2a was positioned 5.5 kb downstream of the 5′ splice site of mouse exon 1, a spacing equivalent to its positioning in human *STMN2* (fig. S6A). Mice with heterozygous or homozygous humanized *Stmn2* gene alleles were viable and developed normally.

Using primary cortical neurons from heterozygous *Stmn2*$^{HumExon2a/+}$ embryos, we confirmed that the presence of human exon 2a and flanking sequences were sufficient to drive altered processing of the humanized *Stmn2* pre-mRNA when TDP-43 was depleted (fig. S6, B and C). An 85% ASO-mediated reduction in TDP-43 level (fig. S6C) triggered nearly complete suppression of full-length mouse *Stmn2* mRNAs from the humanized allele, as indicated by 50% reduction in total *Stmn2* mRNA level in heterozygous cortical neuron cultures (fig. S6D). Usage of cryptic splice-polyadenylation sites was confirmed with abundant, polyadenylated

Stmn2 RNAs containing mouse exon 1 spliced to human exon 2a and whose level was inversely proportional to the level of TDP-43 (fig. S6, C and E). Homozygous humanization of *Stmn2* in a mouse line developing age-dependent motor neuron disease from expression of a disease-linked mutant TDP-43^{Q331K} without loss of nuclear TDP-43 (*57*) neither exacerbated development of disease phenotype (fig. S7, A to D) nor induced cryptic *Stmn2* pre-mRNA misprocessing (fig. S7E) and loss of full-length *Stmn2* mRNA (fig. S7F) in brain or spinal cord by 12 months of age. These findings are consistent with misprocessing of the humanized *Stmn2* pre-mRNA requiring loss of nuclear TDP-43 function.

ALS-FTD model mice that develop cytoplasmic aggregation of TDP-43 from increased expression of wild-type TDP-43 die only days after weaning (*58*, *59*) and do not provide an adequate therapeutic window for ASO-mediated restoration of stathmin-2 levels. To enable in vivo testing of rASOs for their efficacy in restoring stathmin-2 levels within the mammalian nervous system, we generated additional humanized mouse models. Two humanized *Stmn2*^{HumΔGU} founder lines were generated by means of CRISPR-Cas9–mediated genome engineering of mouse ES cells to produce mice carrying a 3.2-kb portion of human *STMN2* intron 1 containing exon 2a (modified to replace the 24-base GU domain that is the TDP-43 binding site with sequence for a 19-base MS2 aptamer) and inserted into the corresponding locus of the first intron of mouse *Stmn2* (Fig. 5A).

In heterozygotes of both lines, the inability of TDP-43 to bind exon 2a^{ΔGU} resulted in constitutive suppression of full-length mouse *Stmn2* mRNAs to 50% of wild-type levels (Fig. 5B) accompanied by chronic use of cryptic splicing and polyadenylation elements encoded by *Stmn2*^{HumΔGU} pre-mRNAs (Fig. 5C) despite normal TDP-43 levels (Fig. 5D). Although *Stmn2*^{HumΔGU} heterozygous mice developed normally and matured without overt neurological symptoms, there was 80% reduced survival to weaning of *Stmn2*$^{HumΔGU/HumΔGU}$ homozygotes identified in heterozygous matings—evidence of developmental lethality driven by *Stmn2* suppression.

Usage of *Stmn2* cryptic splice and polyadenylation sites encoded by the modified human exon 2a was inhibited in the cortex and spinal cord after intracerebral ventricular (ICV) injection of any of three rASOs into cerebral spinal fluid of 2-month-old *Stmn2*$^{HumΔUG/+}$ mice (Fig. 5, E and F). Two weeks after injection, rASO-5 suppressed inclusion of exon 2a by 50 and 35% in the cortex and spinal cord, respectively (Fig. 5, E and F). The suppression of cryptic processing was accompanied by significant restoration of full-length, stathmin-2–encoding mRNAs after injection of two of the three rASOs (rASO-4 and rASO-5) and was maximized by a second administration of rASO-5, yielding 53 and 45% increases of full-length *Stmn2* mRNA production in the cortex and spinal cord and 55 and 40% decreases in truncated *Stmn2* RNAs, respectively (Fig. 5, E to H). The highest stathmin-2 protein restoration in the spinal cord was achieved in mice ICV-injected twice with rASO-5, resulting in almost 80% of its normal expression level in wild-type mice (Fig. 5, I and J). Moreover, there was apparent, dose-dependent stabilization of stathmin-2 protein (Fig. 5, I and J), with the reduced mRNA level in the *Stmn2*$^{HumΔUG/+}$ mice yielding an even larger reduction in stathmin-2 protein, which is evidence supporting a therapeutic threshold for ASO-mediated *STMN2* mRNA restoration inducing larger changes in stathmin-2 protein accumulation.

Discussion

Cytoplasmic accumulation coupled with nuclear clearance of the RNA binding protein TDP-43 is found in affected neurons of ~97% of ALS patients (*10*, *60*, *61*), which suggests that nearly all ALS-causing mechanisms converge on TDP-43 dysfunction. Nuclear loss of TDP-43 is also a common hallmark in approximately half of patients with FTD and Alzheimer's disease (*8*–*11*). Although splicing and 3′ cleavage and polyadenylation of pre-mRNAs are cotranscriptionally coupled (*62*, *63*), by introducing targeted gene editing, we have now determined that TDP-43 binding to the human *STMN2* pre-mRNA sterically blocks recognition of a cryptic 3′ splice site in intron 1, enabling correct pre-mRNA processing and production of a functional, stathmin-2–encoding mRNA. Although therapeutic efforts to target several genetic forms of ALS and FTD are underway (*64*–*66*), direct approaches to restore normal TDP-43 protein localization and function are challenged by many factors, including as yet unidentified mechanisms leading to the initial TDP-43 dysfunction, apparently high sensitivity of healthy neurons to reduction in TDP-43, and a tight autoregulatory mechanism that is proposed to gate TDP-43's own pre-mRNA maturation and translation (*19*, *67*, *68*).

We have identified two approaches to enable restoration of endogenous stathmin-2 accumulation in neurons affected by TDP-43 pathology: targeting dCasRx or cryptic splice blocking ASOs to the *STMN2* pre-mRNA. For the latter approach, we have engineered mice to carry a *Stmn2* gene partially humanized by insertion of the human STMN2 cryptic splice and polyadenylation sequences but without TDP-43 binding. We then demonstrated in vivo proof-of-concept molecular efficacy of ASOs administered into the cerebrospinal fluid in an adult mammalian nervous system to rescue stathmin-2 accumulation to a level that is sufficient to restore axonal regrowth and transport in affected human motor neurons. This approach follows from an initial success in the therapeutic use of a splice-modifying ASO whose action in spinal muscular atrophy restores SMN to motor neurons by correcting missplicing of the pre-mRNA of *SMN2* (*51*, *69*, *70*). Next steps for the development of stathmin-2 restoration as a potential therapy include the development of suitable TDP-43 loss-of-function mouse models with which to determine the phenotypic contribution of *Stmn2* humanization, and direct experimental determination of the functional consequences of stathmin-2 loss from an aging adult mammalian nervous system. Although non-splicing-dependent mechanisms likely play a role in pathophysiology of disease and TDP-43–dependent cryptic splicing affects processing of many mRNA targets [such as *UNC13A* (*21*, *22*)] whose contribution to disease phenotypes remain to be determined, our results provide direct support for ASO-mediated restoration of stathmin-2 as a potential therapeutic approach for ALS and other TDP-43 proteinopathies.

REFERENCES AND NOTES

1. J. K. Nussbacher, R. Tabet, G. W. Yeo, C. Lagier-Tourenne, *Neuron* **102**, 294–320 (2019).
2. C. Lagier-Tourenne, M. Polymenidou, D. W. Cleveland, *Hum. Mol. Genet.* **19** (R1), R46–R64 (2010).
3. A. Nicolas *et al*., *Neuron* **97**, 1268–1283.e6 (2018).
4. K. Weskamp, S. J. Barmada, *Adv. Neurobiol.* **20**, 103–142 (2018).
5. A. D. Gitler, J. Shorter, *Prion* **5**, 179–187 (2011).
6. S. C. Ling, M. Polymenidou, D. W. Cleveland, *Neuron* **79**, 416–438 (2013).
7. J. P. Taylor, R. H. Brown Jr., D. W. Cleveland, *Nature* **539**, 197–206 (2016).
8. K. A. Josephs *et al*., *Acta Neuropathol.* **127**, 811–824 (2014).
9. B. S. Johnson *et al*., *J. Biol. Chem.* **284**, 20329–20339 (2009).
10. T. Arai *et al*., *Biochem. Biophys. Res. Commun.* **351**, 602–611 (2006).
11. C. Amador-Ortiz *et al*., *Ann. Neurol.* **61**, 435–445 (2007).
12. Y. Davidson *et al*., *Acta Neuropathol.* **118**, 359–369 (2009).
13. T. Mishima *et al*., *J. Neuropathol. Exp. Neurol.* **76**, 676–682 (2017).
14. A. C. McKee *et al*., *J. Neuropathol. Exp. Neurol.* **69**, 918–929 (2010).
15. P. T. Nelson *et al*., *Brain* **142**, 1503–1527 (2019).
16. A. L. Nana *et al*., *Acta Neuropathol.* **137**, 27–46 (2019).
17. M. Sun *et al*., *Acta Neuropathol.* **133**, 923–931 (2017).
18. J. R. Tollervey *et al*., *Nat. Neurosci.* **14**, 452–458 (2011).
19. M. Polymenidou *et al*., *Nat. Neurosci.* **14**, 459–468 (2011).
20. C. Lagier-Tourenne *et al*., *Nat. Neurosci.* **15**, 1488–1497 (2012).
21. A.-L. Brown *et al*., *Nature* **603**, 131–137 (2022).
22. X. R. Ma *et al*., *Nature* **603**, 124–130 (2022).
23. J. P. Ling, O. Pletnikova, J. C. Troncoso, P. C. Wong, *Science* **349**, 650–655 (2015).
24. Y. H. Jeong *et al*., *Mol. Neurodegener.* **12**, 13 (2017).
25. Z. Melamed *et al*., *Nat. Neurosci.* **22**, 180–190 (2019).
26. J. R. Klim *et al*., *Nat. Neurosci.* **22**, 167–179 (2019).
27. S. Chauvin, A. Sobel, *Prog. Neurobiol.* **126**, 1–18 (2015).
28. J. E. Shin, S. Geisler, A. DiAntonio, *Exp. Neurol.* **252**, 1–11 (2014).
29. I. Guerra San Juan *et al*., *Neuron* **110**, 1671–1688.e6 (2022).
30. K. L. Krus *et al*., *Cell Rep.* **39**, 111001 (2022).
31. F. Krach *et al*., *Acta Neuropathol.* **136**, 405–423 (2018).
32. M. Prudencio *et al*., *J. Clin. Invest.* **130**, 6080–6092 (2020).
33. L. D. Belmont, T. J. Mitchison, *Cell* **84**, 623–631 (1996).
34. R. Stein, N. Mori, K. Matthews, L. C. Lo, D. J. Anderson, *Neuron* **1**, 463–476 (1988).
35. H. Morii, Y. Shiraishi-Yamaguchi, N. Mori, *J. Neurobiol.* **66**, 1101–1114 (2006).
36. B. M. Riederer *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **94**, 741–745 (1997).
37. J. E. Duncan, N. K. Lytle, A. Zuniga, L. S. Goldstein, *PLOS ONE* **8**, e68324 (2013).
38. E. R. Graf, H. M. Heerssen, C. M. Wright, G. W. Davis, A. DiAntonio, *J. Neurosci.* **31**, 15026–15034 (2011).

39. Y. M. Ayala *et al*., *J. Mol. Biol.* **348**, 575–588 (2005).
40. F. Lim, M. Spingola, D. S. Peabody, *J. Biol. Chem.* **269**, 9006–9010 (1994).
41. G. Rot *et al*., *Cell Rep.* **19**, 1056–1067 (2017).
42. R. Batra *et al*., *Mol. Cell* **56**, 311–322 (2014).
43. H. Liu, L. Tang, *Curr. Genomics* **14**, 49–55 (2013).
44. A. Masuda *et al*., *Genes Dev.* **29**, 1045–1057 (2015).
45. G. Martin, A. R. Gruber, W. Keller, M. Zavolan, *Cell Rep.* **1**, 753–763 (2012).
46. D. C. Di Giammartino, K. Nishida, J. L. Manley, *Mol. Cell* **43**, 853–866 (2011).
47. B. R. So *et al*., *Mol. Cell* **76**, 590–599.e4 (2019).
48. R. Elkon, A. P. Ugalde, R. Agami, *Nat. Rev. Genet.* **14**, 496–506 (2013).
49. B. Tian, J. H. Graber, *Wiley Interdiscip. Rev. RNA* **3**, 385–396 (2012).
50. S. Konermann *et al*., *Cell* **173**, 665–676.e14 (2018).
51. R. S. Finkel *et al*., *Lancet* **388**, 3017–3026 (2016).
52. C. F. Bennett, H. B. Kordasiewicz, D. W. Cleveland, *Annu. Rev. Pharmacol. Toxicol.* **61**, 831–852 (2021).
53. N. H. Alami *et al*., *Neuron* **81**, 536–543 (2014).
54. P. Guedes-Dias, E. L. F. Holzbaur, *Science* **366**, eaaw9997 (2019).
55. E. Perlson, S. Maday, M. M. Fu, A. J. Moughamian, E. L. Holzbaur, *Trends Neurosci.* **33**, 335–344 (2010).
56. J. N. Sleigh *et al*., *Cell Rep.* **30**, 3655–3662.e2 (2020).
57. E. S. Arnold *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **110**, E736–E745 (2013).
58. H. Wils *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3858–3863 (2010).
59. L. A. Becker *et al*., *Nature* **544**, 367–371 (2017).
60. I. R. Mackenzie *et al*., *Ann. Neurol.* **61**, 427–434 (2007).
61. M. Neumann *et al*., *Science* **314**, 130–133 (2006).
62. F. Rigo, H. G. Martinson, *Mol. Cell. Biol.* **28**, 849–862 (2008).
63. A. Kyburz, A. Friedlein, H. Langen, W. Keller, *Mol. Cell* **23**, 195–205 (2006).
64. T. Miller *et al*., *N. Engl. J. Med.* **383**, 109–119 (2020).
65. V. A. Korobeynikov, A. K. Lyashchenko, B. Blanco-Redondo, P. Jafar-Nejad, N. A. Shneider, *Nat. Med.* **28**, 104–116 (2022).
66. B. D. Boros, K. M. Schoch, C. J. Kreple, T. M. Miller, *Neurotherapeutics* **19**, 1145–1158 (2022).
67. Y. M. Ayala *et al*., *EMBO J.* **30**, 277–288 (2011).
68. S. E. Avendaño-Vázquez *et al*., *Genes Dev.* **26**, 1679–1684 (2012).
69. Y. Hua, T. A. Vickers, H. L. Okunola, C. F. Bennett, A. R. Krainer, *Am. J. Hum. Genet.* **82**, 834–848 (2008).
70. C. F. Bennett, A. R. Krainer, D. W. Cleveland, *Annu. Rev. Neurosci.* **42**, 385–406 (2019).

ACKNOWLEDGMENTS

The generation and development of the CRISPR-Cas9 mutant mouse models was supported by the Jackson Laboratory Genome Engineering Technology Core. **Funding:** This work was supported by NINDS/NIH R01NS112503 (to D.W.C. and C.L.-T.); NINDS/NIH RF1NS124203 (D.W.C., C.L.-T., and C.L.); ALS Finding a Cure (C.L.-T.); The Massachusetts Center for Alzheimer Therapeutic Science (C.L.-T.); The Sean M. Healey & AMG Center for ALS at Mass General (C.L.-T.); U42 Mutant Mouse Resource Research Center OD010921 (C.L.); Ruth Kirschstein Institutional National Research Service Award T32 GM008666. (M.W.B. and M.S.B.), and T32 AG 66596-2 (M.S.B.); The Packard Center for ALS Research (D.W.C. and M.W.B.); The ALS association (D.W.C., M.W.B., and S.V.-S.); MDA development grants (Z.M. and J.L.-E.); The BrightFocus Foundation (A.R.A.d.A.Q.); and Cancer Center Support Grant CA034196 to The Jackson Laboratory. **Author contributions:** Conceptualization: M.W.B., Z.M., J.L.-E., A.Z., C.L., F.R., K.L., P.J.-n., C.F.B., C.L.-T., and D.W.C. Experiments: Z.M., M.W.B., J.L.-E., M.S.B., K.L., A.Z., M.P., E.G.-G., R.M., S.V.-S., S.C., M.B.-H., V.T., J.W.A., E.A., I.S.N., and A.R.A.d.A.Q. Analysis: M.W.B., Z.M., J.L.-E., K.L., P.J.-n., F.R., C.F.B., M.P., E.G.-G., S.M., A.Z., C.L., C.L.-T., and D.W.C. Key methodology and resources: K.L., P.J.-n., F.R., C.F.B., M.P., A.Z., and C.L. Writing: Z.M., M.W.B., C.L.-T., and D.W.C. **Competing interests:** A patent application has been filed describing composition and methods for restoring stathmin-2 expression, publication number US 2021-0252039. D.W.C. and Z.M. are inventors on the patent. C.F.B., K.L., P.J.-n., and F.R. are employees of Ionis Pharmaceuticals. D.W.C. is a consultant for Ionis Pharmaceuticals. **Data and materials availability:** SH-SY5Y and N2A cells were obtained for research use through a materials transfer agreement with ATCC. Plasmids were obtained through a materials transfer agreement with Addgene. All mouse strains have been made publicly available through The Jackson Laboratory. Other materials are available under materials transfer agreement for noncommercial replication or extension of this work, upon request to corresponding authors.

Materials and methods and references cited therein can be found in the main article online.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq5622
Materials and Methods
Figs. S1 to S7
Table S1
MDAR Reproducibility Checklist

Submitted 17 April 2022; resubmitted 2 December 2022
Accepted 2 February 2023
10.1126/science.abq5622

CELL BIOLOGY

Bacteria require phase separation for fitness in the mammalian gut

Emilia Krypotou[1,2], Guy E. Townsend II[1,2,3], Xiaohui Gao[1], Shoichi Tachiyama[1,2], Jun Liu[1,2], Nick D. Pokorzynski[1], Andrew L. Goodman[1,2], Eduardo A. Groisman[1,2]*

Therapeutic manipulation of the gut microbiota holds great potential for human health. The mechanisms bacteria use to colonize the gut therefore present valuable targets for clinical intervention. We now report that bacteria use phase separation to enhance fitness in the mammalian gut. We establish that the intrinsically disordered region (IDR) of the broadly and highly conserved transcription termination factor Rho is necessary and sufficient for phase separation in vivo and in vitro in the human commensal *Bacteroides thetaiotaomicron*. Phase separation increases transcription termination by Rho in an IDR-dependent manner. Moreover, the IDR is critical for gene regulation in the gut. Our findings expose phase separation as vital for host-commensal bacteria interactions and relevant for novel clinical applications.

The gut microbiota plays a critical role in human health, with some species promoting wellbeing and others causing various diseases (*1*). Manipulating the gut microbiota holds excellent clinical promise but requires identifying factors and mechanisms that enable beneficial bacteria to colonize the gut. One of the most abundant bacterial species in the human gut, *Bacteroides thetaiotaomicron*, is associated with lean and healthy individuals (*2*) and is being tested in clinical trials as a biotherapeutic for gastrointestinal disorders (*3*). We therefore investigated what controls *B. thetaiotaomicron*'s ability to colonize the mammalian gut.

We establish that *B. thetaiotaomicron*'s fitness in the murine gut is facilitated by a unique domain of its transcription termination factor Rho that is necessary for liquid-liquid phase separation (LLPS). LLPS is a phenomenon guided by protein-protein or protein-RNA electrostatic interactions and results in the formation of biomolecular condensates (*4*, *5*). These condensates enable cells to perform biological processes often related to RNA regulation in distinct, membraneless subcellular compartments (*6*–*8*). In eukaryotes, LLPS plays fundamental physiological roles, including (i) arresting translation of inactivated ribosomes (*9*), (ii) controlling transcription rate by condensing transcription factors with nascent RNA (*10*), and (iii) storing untranslated mRNAs and translation factors stalled at initiation codons triggered by stress conditions (stress granules) (*9*). In bacteria, LLPS regulates several processes (*11*–*13*), including (i) mRNA decay in RNA degradosome condensates (BR-bodies) of *Caulobacter crescentus* (*14*) and (ii) RNA polymerase (RNAP) clustering during exponential growth of *Escherichia coli* (*15*).

Rho is an adenosine triphosphate (ATP)-dependent RNA helicase essential in gram-negative bacteria, including *B. thetaiotaomicron* (*16*, *17*). Rho controls gene expression by terminating transcription at the end of genes or within mRNA leaders, thereby modifying the RNA abundance of the downstream regions in response to specific signals (*18*, *19*). Rho is highly conserved but harbors an additional domain in several species that varies in sequence, length, and properties (*20*) (fig. S1). For example, the additional domain of *Clostridium botulinum* Rho exhibits prion properties when expressed in *E. coli* (*21*), and the additional domains of the *Bacteroides fragilis* and *Micrococcus*

[1]Department of Microbial Pathogenesis, Yale School of Medicine, 295 Congress Avenue, New Haven, CT 06536, USA. [2]Yale Microbial Sciences Institute, P.O. Box 27389, West Haven, CT 06516, USA. [3]Department of Biochemistry and Molecular Biology, Penn State College of Medicine, 700 HMC Crescent Road, Hershey, PA 17033, USA.
***Corresponding author. Email: eduardo.groisman@yale.edu**

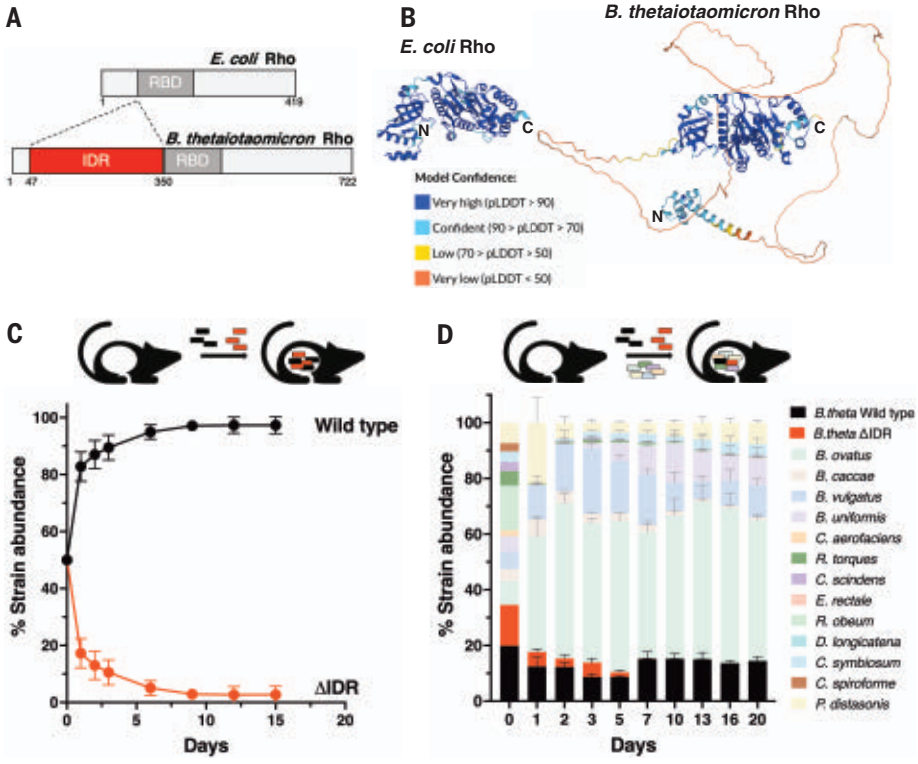


Fig. 1. *B. thetaiotaomicron* Rho harbors an intrinsically disordered region required for fitness in the murine gut. (**A**) The *B. thetaiotaomicron* Rho protein harbors a 303 amino acid–long domain absent from well-characterized homologs, such as that from *E. coli*. Predicted to be intrinsically disordered (IDR), the identified domain is immediately adjacent to Rho's RNA binding domain (RBD). (**B**) AlphaFold (*60*, *61*) prediction of the structures of the *E. coli* and *B. thetaiotaomicron* Rho proteins (source UniProt) highlighting the IDR in the latter. The N- and C-termini of the structures are indicated with N and C, respectively. (**C**) Relative abundance of isogenic *B. thetaiotaomicron* strains expressing WT (GT1504) or ΔIDR (GT1506) Rho at the indicated times in the gut of germ-free mice (n = 5 Swiss Webster mice); the strains were present in a 1:1 ratio in the inoculum. (**D**) Relative abundance of isogenic *B. thetaiotaomicron* strains harboring WT (AK310) or ΔIDR (AK312) Rho and 13 species representing the major phyla in the human gut during gut colonization in germ-free mice (n = 5 C57BL/6 mice); the *B. thetaiotaomicron* strains were supplied at ~1:1 ratio (2×10^8 CFU WT Rho- versus 1.5×10^8 CFU ΔIDR Rho-expressing strains) in the inoculum. For (C) and (D), bacterial abundance was measured by qPCR of genomic DNA recovered from mouse fecal samples over time. SD error bars are shown.

luteus Rho proteins mediate RNA interactions during in vitro transcription termination assays (*22*, *23*). However, the physiological role(s) that the additional Rho domain plays in native bacteria has remained unknown.

The *B. thetaiotaomicron* Rho intrinsically disordered region (IDR) is required for fitness in the gut

The 722-residue–long Rho from *B. thetaiotaomicron* harbors a 303-residue region near its N terminus not found in the 419-residue Rho proteins from the well-studied bacteria *E. coli* and *Salmonella enterica* serovar Typhimurium (*Salmonella*) and predicted to be intrinsically disordered (IDR) based on amino acid content and sequence (MobiDB: database of protein disorder and mobility annotations) (*22*) (Fig. 1, A and B). Rho proteins from several other bacterial species harbor an additional domain at the same position as the *B. thetaiotaomicron* Rho (*20*). However, these domains differ from the 303-residue region of the *B. thetaiotaomicron* Rho in amino acid sequence and length [fig. S1 and (*20*)], suggesting that the additional domain was acquired independently during evolution. The additional domains of closely related species share amino acid content but not sequence. Within the Bacteroidetes, the additional domain comprises a Lys- and Glu-rich region and an Asn- and Gln-rich region separated by a short, highly conserved amino acid sequence (fig. S1).

To establish the role that the Rho IDR itself plays in *B. thetaiotaomicron* physiology, we engineered isogenic *B. thetaiotaomicron* strains deleted for the native *rho* gene and expressing wild-type (WT) or ΔIDR *rho* from the chromosomal site for the insertion element NBU2 (fig. S2A). The Δ*rho* strain expressing Δ*IDR rho* (ΔIDR strain) was viable under laboratory conditions and exhibited slightly faster growth during exponential phase in minimal medium supplemented with glucose (fig. S3A) and slightly higher final yield in minimal and rich media than the isogenic strain expressing WT *rho* (fig. S3B). By contrast, the ΔIDR strain was readily outcompeted in the gut of germ-free mice by the WT Rho-expressing strain, which accounted for 95% of the bacterial population by day 6 (Fig. 1C). Because removal of the IDR did not affect Rho protein abundance (fig. S4), these results suggest that *B. thetaiotaomicron* requires the Rho IDR for fitness in the mammalian gut.

Gut colonization experiments using a different set of isogenic strains that expressed WT or ΔIDR *rho* from the native genetic locus (fig. S2B) and a different germ-free mouse strain resulted in the same outcome: the ΔIDR strain was readily outcompeted by the wild type despite the ΔIDR strain abundance in the inoculum being higher than that of WT *B. thetaiotaomicron* (fig. S2C). Moreover, the ΔIDR strain was also outcompeted by the WT strain in ex–germ-free mice harboring a complex bacterial community comprising 13 bacterial species representative of the main phyla in the human gut (Fig. 1D). Cumulatively, these results robustly demonstrate that Rho's IDR is essential for *B. thetaiotaomicron* fitness in the murine gut.

The IDR promotes Rho phase separation in vitro

To investigate how IDR furthers gut colonization, we examined whether the Rho phase separates because intrinsically disordered regions of RNA-binding proteins often promote LLPS (*5*–*8*). Differential interference contrast (DIC) microscopy revealed that purified WT Rho protein forms droplets in vitro at concentrations >1 μM in physiological salt conditions (150 mM KCl) (*24*) (Fig. 2A and fig. S5A). By contrast, no droplets were detected for the ΔIDR protein even at a concentration of 20 μM (Fig. 2A). Because the Rho intracellular concentration is ~1 μM (see Materials and Methods), the in vitro LLPS occurs at a physiological protein concentration.

The size and abundance of the droplets formed by the WT Rho protein are dependent on the salt concentration, both increasing at lower concentrations (fig. S6, A and B). Addition of a total RNA extract from *B. thetaiotaomicron* resulted in larger and more abundant droplets for WT Rho but had no effect on the ΔIDR protein (Fig. 2B and fig. S5B), in agreement with the notion that RNAs often stimulate LLPS in RNA-binding proteins harboring disordered domains (*4*, *25*, *26*). Droplets formed by WT Rho underwent fusion (Fig. 2C) and dissolved in the presence of LLPS inhibitor 1,6-hexanediol but not the control 2,5-hexanediol (*27*, *28*) (fig. S6C). A fluorescently tagged WT Rho protein (Rho_mNeonGreen) behaved similarly to untagged Rho, forming

droplets that responded to salt, protein, and RNA concentrations and were dissolved by 1,6-hexanediol but not the control 2,5-hexanediol (fig. S7), as reported for other RNA-binding, phase-separating proteins (*4*, *25*–*28*). Fluorescence recovery after photobleaching (FRAP) experiments using droplets formed by Rho_mNeonGreen and RNA exhibited slow fluorescence recovery (Fig. 2, E and F), consistent with droplets formed by LLPS with gel-like properties. These results establish that the *B. thetaiotaomicron* Rho protein exhibits IDR-dependent LLPS in vitro.

The purified IDR domain also formed droplets but only in the presence of the RNA extract (Fig. 2B and fig. S5B). Fluorescently labeled RNA localized within droplets formed by Rho or the IDR (Fig. 2D), suggesting that the negatively charged RNA promotes LLPS through electrostatic interactions with the positively charged IDR (fig. S1), as reported for other proteins displaying LLPS (*25*, *26*). We established that low RNA:Rho ratios promote droplet formation, whereas high ratios hinder it (Fig. 2G and fig. S5C) (*4*). By contrast, droplets formed by the IDR increased substantially at high RNA concentrations independently of the protein concentration (fig. S8). These data establish that: (i) the IDR is necessary and sufficient for in vitro LLPS; (ii) RNA abundance regulates droplet formation through the IDR; and (iii) a region of Rho outside the IDR modulates its LLPS properties.

The IDR promotes Rho phase separation in vivo

We examined IDR's ability to promote Rho LLPS in vivo by investigating *B. thetaiotaomicron* experiencing carbon starvation, a condition inducing or activating transcription factors required for gut colonization (*29*, *30*), or carbon-replete conditions. Immunofluorescence revealed that HA-tagged WT Rho and ΔIDR Rho localized throughout the cytoplasm (Fig. 3A). However, WT Rho displayed patchy dispersion and formed clusters that increased under carbon starvation and disappeared following 1,6-hexanediol treatment, whereas ΔIDR Rho was evenly distributed under all conditions (Fig. 3A).

Rho phase separates while *B. thetaiotaomicron* is in the murine gut because distinct foci were observed in bacteria harvested from the cecal contents of germ-free mice monocolonized with *B. thetaiotaomicron* expressing WT Rho but not with the isogenic ΔIDR Rho mutant (Fig. 3B). Cryo–electron tomography (cryo-ET) revealed high-density circular formations in WT *B. thetaiotaomicron* harvested from the gut but not in the isogenic ΔIDR *rho* mutant (Fig. 3B and movies S1 and S2). While the circular formations are IDR-dependent, proteins other than WT Rho and/or RNAs may be part of these assemblies. Further investigations will explore the relationship between the high-density circular formations observed

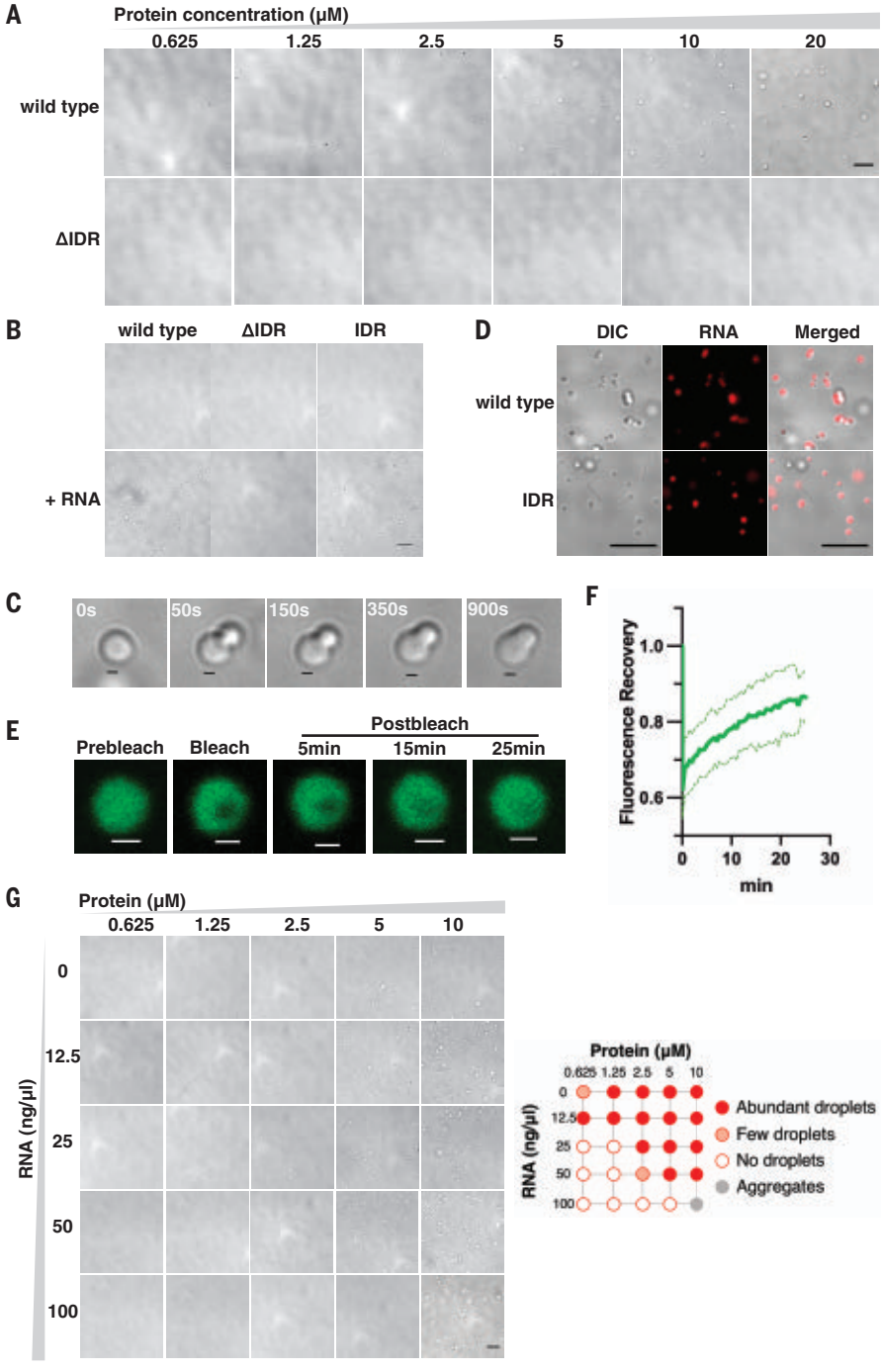


Fig. 2. Rho exhibits IDR-dependent phase separation in vitro. (**A**) DIC microscopy of WT Rho and ΔIDR Rho proteins at the indicated protein concentrations reveals droplet formation by the former but not by the latter. (**B**) DIC microscopy of WT Rho, ΔIDR Rho, and IDR proteins (2.5 μM) in the presence of total *B. thetaiotaomicron* RNA extract (25 ng/μl). (**C**) Time-lapse DIC microscopy of WT Rho (10 mM) droplet fusion in the presence of total *B. thetaiotaomicron* RNA extract (50 ng/ml). (**D**) DIC and fluorescence microscopy of WT (5 mM) and IDR (2.5 mM) proteins with fluorescently labeled (TM-rhodamine) total *B. thetaiotaomicron* RNA extract (12.5 ng/ml and 100 ng/ml, respectively). (**E**) FRAP of WT Rho_mNeonGreen (2.5 μM) with total *B. thetaiotaomicron* RNA extract (12.5 ng/μl) at 50 mM KCl. (**F**) Normalized fluorescence recovery of the FRAP experiment in (E). Mean and SD (dashed lines) of n = 28 droplets are shown. (**G**) (Left) DIC microscopy of WT Rho protein at the indicated concentrations in the presence of total *B. thetaiotaomicron* RNA extract at the indicated concentrations. (Right) Schematic of droplet formation data shown in left panel. Experiments were carried out in the presence of 150 mM KCl unless indicated otherwise. n = 3 independent experiments. Scale bars, (A), (B), (C), and (G) 10 μm; (D) and (E) 1 μm.

by cryo-ET and the foci observed by immunofluorescence in the strain expressing WT Rho (Fig. 3B).

Quantification of Rho clustering using the dispersion of the immunofluorescence signal (*15*) (see Materials and Methods and fig. S9) revealed significantly higher clustering for WT Rho than for ΔIDR Rho when bacteria were grown in glucose, subjected to carbon starvation conditions, or harvested from the mouse gut (Fig. 3, C and D, and fig. S10). 1,6-hexanediol treatment decreased clustering values for bacteria expressing WT Rho to those of bacteria expressing ΔIDR Rho, which were not affected by the treatment (Fig. 3D). The transcription factor BT4338 (*30*)—which served as a negative control because it lacks a disordered domain and was not expected to exhibit LLPS—behaved similarly to the ΔIDR Rho. That is, BT4338 was evenly localized throughout the cytoplasm, not affected by 1,6-hexanediol treatment, and displayed similar clustering under all investigated conditions (Fig. 3, A and D).

When expressed in trans in a Δ*IDR rho* mutant, the IDR formed cytoplasmic puncta and displayed high clustering that was reduced upon 1,6-hexanediol treatment (Fig. 3, A and D). The IDR expressed in trans clustered similarly in bacteria grown in glucose or experiencing carbon starvation, consistent with the in vitro LLPS experiments demonstrating that the purified IDR domain behaves differently from the purified full-length WT Rho protein (Fig. 2 and fig. S8). The results presented above indicate that Rho forms condensates in an IDR-dependent manner in *B. thetaiotaomicron* and that the number of condensates increases upon carbon starvation and in the murine gut.

IDR-dependent phase separation governs transcription termination by Rho

How do LLPS and the IDR control Rho's ability to terminate transcription? We designed in vitro transcription termination assays to determine whether LLPS affects Rho-dependent transcription termination efficiency and/or specificity and whether the IDR modifies Rho-dependent termination under non-LLPS conditions. Initially, we used the *Salmonella mgtA* leader as a template because it has an established Rho-dependent terminator (*31*), together with *Salmonella* RNAP, sigma70, and NusG proteins and also because the *mgtA* RNA does not promote Rho LLPS in vitro (Fig. 4A). We determined that *B. thetaiotaomicron* Rho requires NusG to terminate transcription in the *mgtA* leader and produces the same termination product as the *Salmonella* Rho protein used as control (fig. S11) (*31*), demonstrating that *B. thetaiotaomicron* Rho is functional in a heterologous system. We then carried out experiments using a salt concentration that promotes LLPS (100 mM KCl) and one that does not (200 mM KCl) (Fig. 4B and fig. S12).

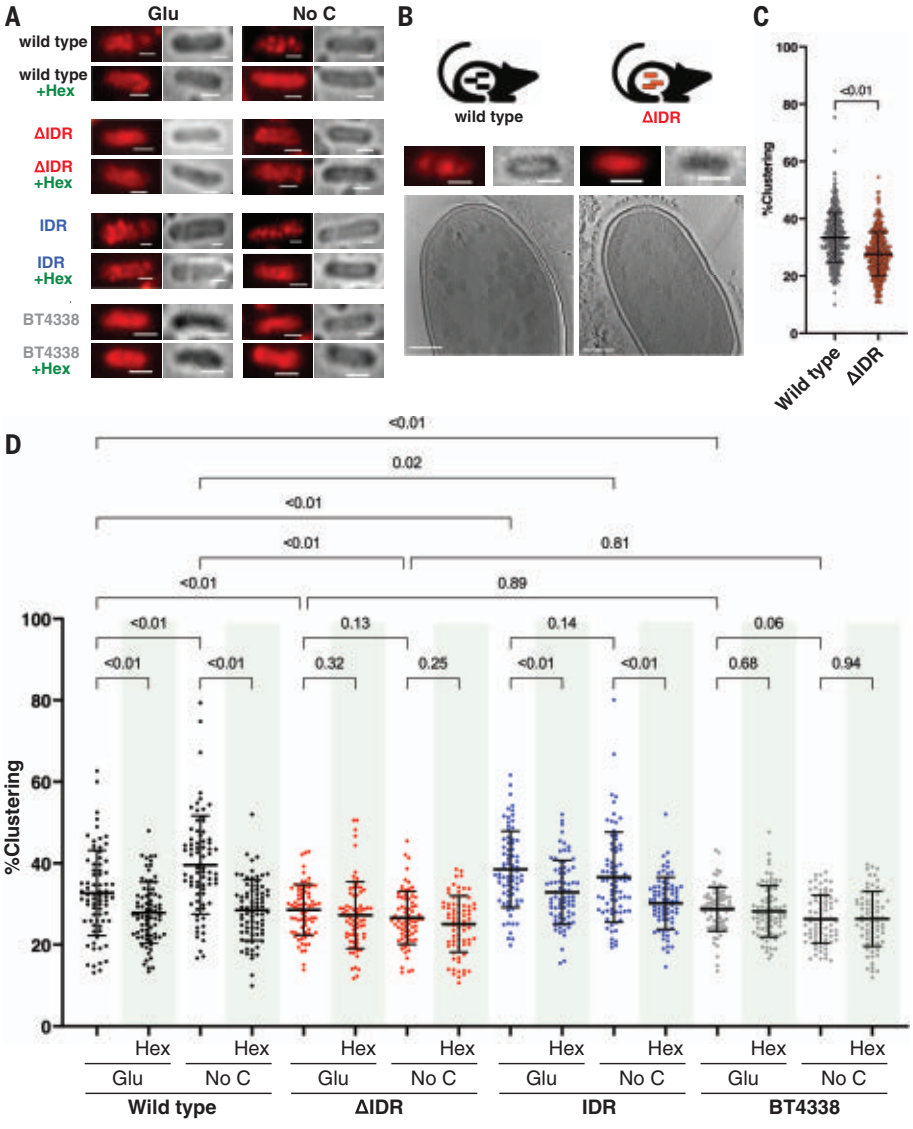


Fig. 3. Rho exhibits IDR-dependent phase separation in vivo. (**A**) Fluorescence microscopy of immunostained isogenic *B. thetaiotaomicron* strains expressing HA-tagged WT Rho (AK82), ΔIDR Rho (AK86), IDR (AK393), or BT4338 (GT1481). Bacteria were grown until mid-exponential phase in minimal media with glucose (Glu), then subjected to 30 min carbon starvation (No C) and/or 5 min 5% 1,6-hexanediol (Hex). (**B**) Fluorescence microscopy and 3D reconstructions using cryo-ET of bacteria harvested from the gut of germ-free mice monocolonized with *B. thetaiotaomicron* expressing HA-tagged WT Rho (AK82) or ΔIDR Rho (AK86). *n* = 3 mice per strain. (**C**) Quantification of protein clustering data shown in (B). (Data points represent clustering values of individual cells from three independent experiments; *n* = 405). (**D**) Quantification of protein clustering data shown in (A). (Data points represent clustering values of individual cells from three independent experiments; *n* = 75). Scale bars: Immunofluorescence in (A and B) 1 μm and cryo-ET (B) 200 nm. *P* values: unpaired *t* test in (C); Fisher's Least Significant Difference (LSD) test was performed only for depicted pairwise comparisons (D).

Under LLPS-promoting conditions, WT Rho displayed significantly higher transcription termination activity within the *mgtA* leader than ΔIDR Rho (Fig. 4, C and D). By contrast, WT and ΔIDR Rho exhibited similarly low termination activity under non-LLPS conditions (Fig. 4, C and D). The high-salt non-LLPS conditions slightly decreased RNAP activity as previously reported (*32*). In addition, termination by ΔIDR Rho was lower at high salt than at low salt, indicative of salt also exerting an effect independently of LLPS. Control experiments revealed that the Rho-specific inhibitor bicyclomycin (BIC) (*33*) prevents termination by WT and ΔIDR Rho under both LLPS and non-LLPS conditions (Fig. 4, C and D). In sum, under LLPS conditions, WT Rho has higher termination activity on the *mgtA* template

than ΔIDR Rho. These results suggest that the reported higher termination displayed by WT *B. fragilis* Rho on a different template relative to the ΔIDR derivative in low-salt media (*22*) is likely the result of LLPS.

To further explore the role of LLPS in Rho-dependent transcription termination, we used a template corresponding to the leader region and part of the coding region of the *B. thetaiotaomicron roc* (*BT3172*) gene (*34*). We chose this template because *roc* mRNA amounts increased ~10 times upon BIC treatment in vivo (fig. S13) and because the *roc* RNA promoted Rho LLPS in vitro, in contrast to the *mgtA* RNA (Fig. 4A). Under LLPS-promoting conditions,

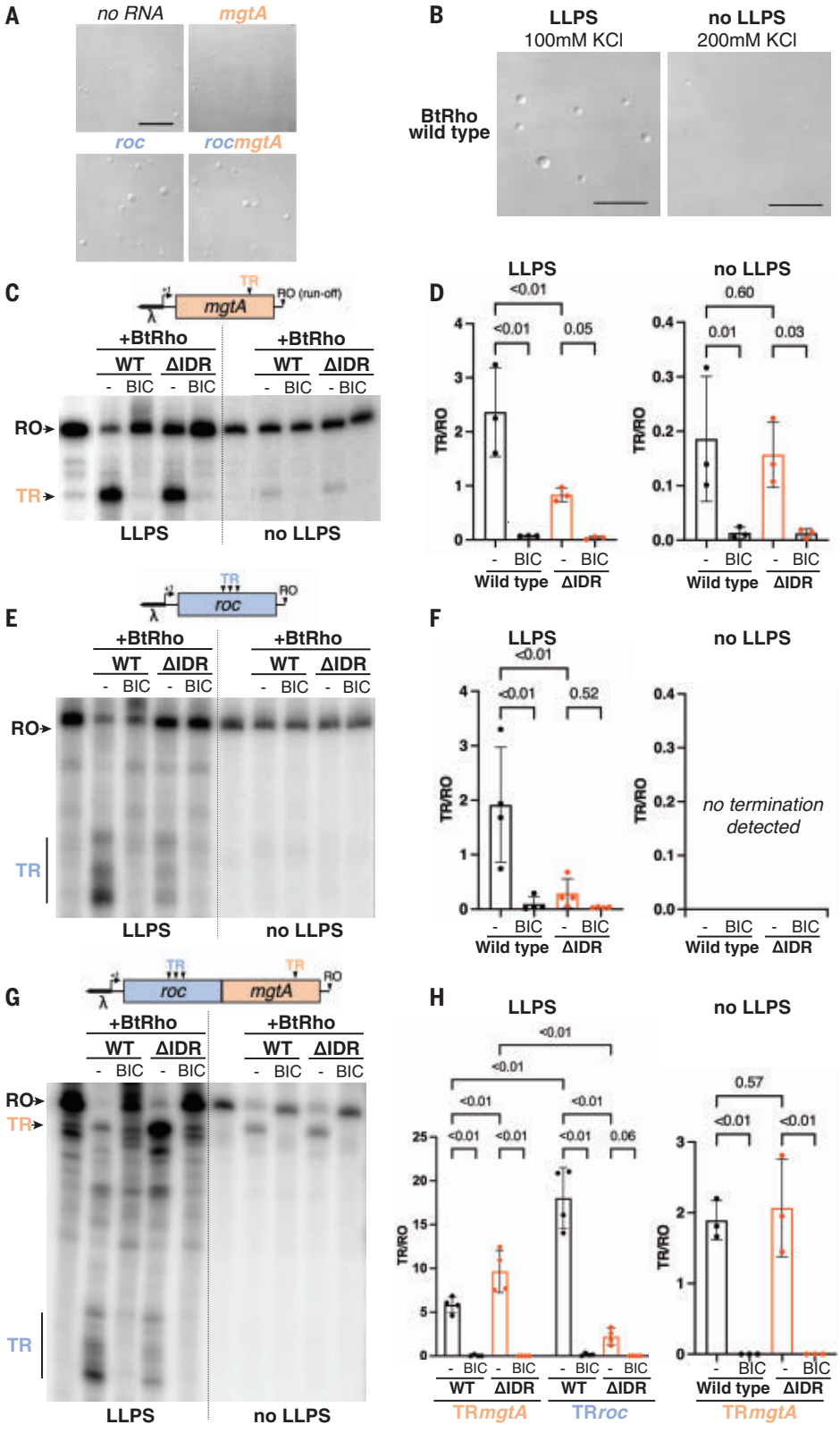


Fig. 4. IDR-dependent LLPS control of Rho transcription termination in vitro. (**A**) DIC microscopy of WT Rho protein (2.5 μM) in the presence of 50 ng/μl of *mgtA* RNA, *roc* RNA, or *rocmgtA* RNA, and 150 mM KCl. (**B**) DIC microscopy of WT Rho protein used in the in vitro transcription assays under LLPS-promoting (100 mM KCl) or non–LLPS-promoting (200 mM KCl) conditions. In vitro transcription of DNA templates corresponding to *mgtA* (**C**), *roc* (**E**), and *rocmgtA* (**G**) in the presence of WT Rho or ΔIDR Rho ± Rho inhibitor bicyclomycin (BIC) under LLPS-promoting (100 mM KCl) or non–LLPS-promoting (200 mM KCl) conditions. Arrows indicate runoff (RO) and termination (TR) products. (**D**, **F**, and **H**) Relative transcript abundance (termination/run off) for data presented in (C), (E), and (G), respectively. Scale bar: 10 μm. 1 μM Rho protein was used in (B) to (H). Transcription from the three templates was driven by the λp_R promoter. NusG (300 nM) was included in all transcription assays. *P* values: Fisher's LSD test was performed only for depicted pairwise comparisons, $n \geq 3$ independent experiments, SD error bars are shown. Bands shown in (C) are from the same gel, but during preparation of this figure two lanes between lanes 1 and 2 and one lane between lanes 6 and 7 were cropped from the original image (see fig. S16 for uncropped image).

WT Rho produced multiple termination products with high efficiency (Fig. 4, E and F), typical of Rho-dependent terminators (*35*), whereas ΔIDR Rho showed minimal termination activity (Fig. 4, E and F). As expected, BIC inhibited the appearance of termination products (Fig. 4, E and F). By contrast, no termination products were detected for either Rho protein under non-LLPS conditions (Fig. 4, E and F). Therefore, efficient termination in the *roc* template requires IDR-dependent LLPS of WT Rho.

Next, we examined how LLPS affects transcription termination by WT and ΔIDR Rho proteins on an engineered template consisting of the *roc* sequence at the 5′ end and the *mgtA* sequence at the 3′ end (Fig. 4G). This template harbors sites terminated by Rho with different efficiencies under non-LLPS (*mgtA*, high; *roc*, low) and can elicit LLPS (Fig. 4A). Under LLPS-promoting conditions, WT Rho promoted termination primarily within *roc*, whereas the ΔIDR Rho promoted termination primarily within *mgtA* (Fig. 4, G and H). That ΔIDR Rho appears to be more active than WT Rho in terminating at the downstream *mgtA* reflects that WT Rho promotes termination at the upstream *roc* (Fig. 4, G and H). By contrast, under non-LLPS conditions, WT and ΔIDR Rho promoted termination only within *mgtA* (Fig. 4, G and H), which is higher in the *roc-mgtA* template (Fig. 4, G and H) than in the *mgtA* template (Fig. 4, C and D). (The different termination within *mgtA* observed with the *mgtA* and *rocmgtA* templates may reflect differences in the secondary structures of the *mgtA* and *rocmgtA* RNAs affecting Rho access to its binding sites.) The IDR itself does not appear to contribute to Rho's termination activity under non-LLPS conditions because WT and ΔIDR Rho exhibited similarly robust termination activity under these conditions (Fig. 4, G and H). We conclude that LLPS increases the transcription termination activity of WT Rho and that certain RNAs (e.g., *roc*) are terminated primarily under LLPS-promoting conditions.

Central metabolism genes show IDR-dependent expression in the gut

To understand how LLPS control of Rho transcription termination affects *B. thetaiotaomicron* gene expression in the gut, we determined the bacterial mRNA abundance in cecal contents of ex–germ-free mice monocolonized with isogenic *B. thetaiotaomicron* strains expressing either WT Rho or ΔIDR Rho. RNA was harvested from bacteria in mice sacrificed 3 days after gavage when both strains had reached similar numbers (fig. S14).

The mRNA abundance of ~400 genes differed between the two strains (FDR-adjusted *P* value <0.05): 208 of the genes were present in greater than twofold lower amounts and

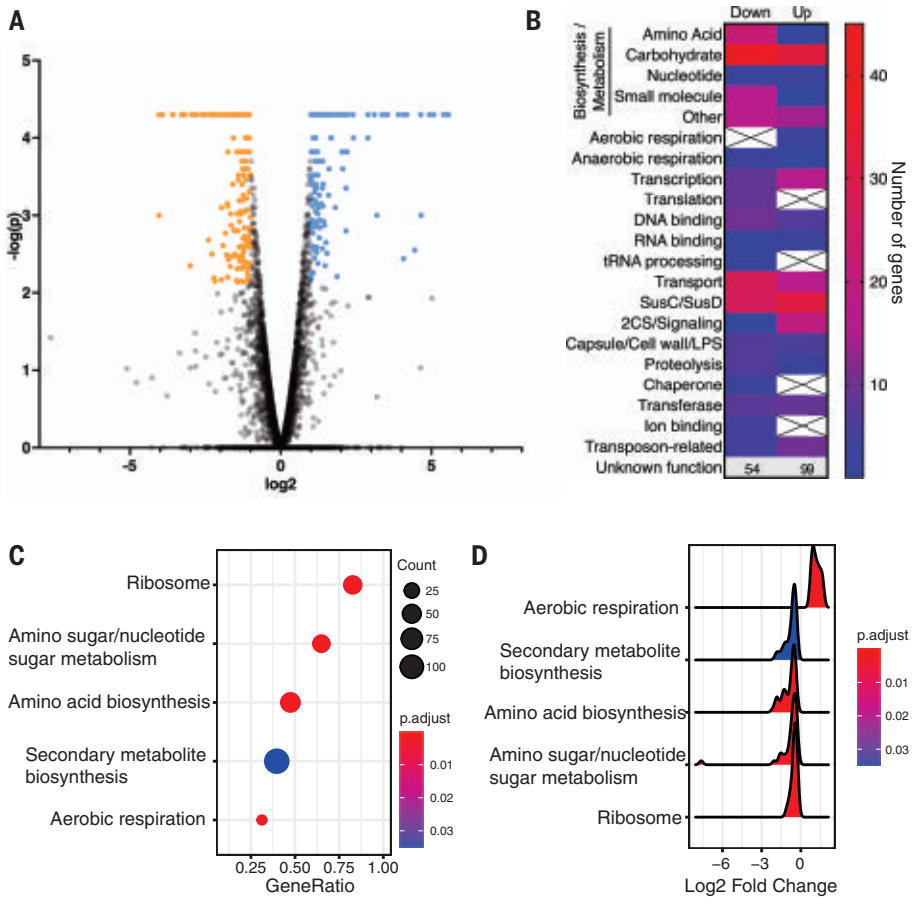


Fig. 5. *B. thetaiotaomicron* exhibits IDR-dependent gene expression in the murine gut. RNA-seq analysis was performed in cecal contents of ex–germ-free mice (C57BL/6 mice, *n* = 5) monocolonized with strains harboring WT Rho (AK310) or ΔIDR (AK312) Rho. (**A**) Volcano plot of gene RNA abundance in the WT versus ΔIDR Rho strains as log2-fold change versus -log(p). Genes >2-fold up-regulated (blue) or down-regulated (orange) in the ΔIDR background are highlighted. (FDR-adjusted *P* value <0.05). (**B**) Heat map of genes differentially expressed in strains harboring WT Rho versus ΔIDR Rho based on molecular function. (**C**) KEGG pathway enrichment analysis of differentially expressed in strains harboring WT Rho versus ΔIDR Rho. Identified pathways with enrichment p-value <0.05 and number of genes per pathway shown. (**D**) Log2 fold change of genes identified in (C).

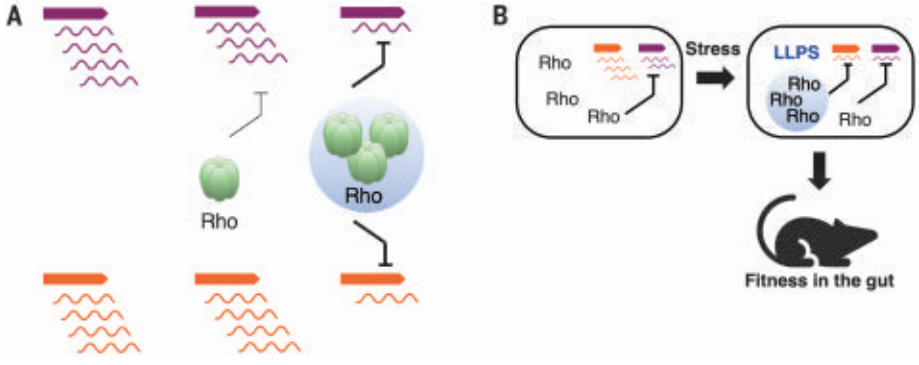


Fig. 6. LLPS of *B. thetaiotaomicron* Rho governs *B. thetaiotaomicron* gene expression in the gut. (**A**) Conditions that elicit IDR-dependent LLPS of Rho (highlighted with a blue circle) increase transcription termination by WT Rho. Some RNAs can be terminated with low efficiency even without LLPS (i.e., *mgtA*, purple schematic) but other RNAs require LLPS for termination (i.e., *roc*, orange schematic). (**B**) When *B. thetaiotaomicron* experiences specific stress conditions, Rho exhibits LLPS in an IDR-dependent manner, which increases Rho-dependent transcription termination thereby altering global gene expression in ways that further *B. thetaiotaomicron* fitness in the gut.

185 in greater than twofold higher amounts in the ΔIDR *rho* mutant compared with the isogenic WT strain (Fig. 5A). Numerous genes involved in central cellular pathways, including those of amino acid and small molecule biosynthesis, tRNA processing, and protein translation and folding, were down-regulated in the ΔIDR strain, whereas genes related to transcription and two-component system signaling were up-regulated. Other genes exhibiting IDR-regulated expression are involved in carbohydrate metabolism and carbohydrate acquisition by SusC- or SusD-like proteins (*36*), which mediate binding and uptake of complex polysaccharides (Fig. 5B). Importantly, several IDR-regulated genes are required for bacterial fitness in the gut, including those involved in acquisition of the essential vitamin B12, which *B. thetaiotaomicron* is unable to synthesize and secures from the host diet (*17*, *37*) (table S1).

KEGG pathway enrichment analysis revealed that the ΔIDR *rho* strain has lower mRNA abundance of genes participating in translation and biosynthetic pathways for amino acids, nucleotides, and secondary metabolites but higher abundance of those participating in aerobic respiration than the WT strain (Fig. 5, C and D, and table S1). Notably, the mRNA amounts of anaerobic respiration genes—which have been implicated in gut colonization in several *Bacteroides* species (*25*, *26*)—were lower in the ΔIDR *rho* mutant than in the WT strain (table S1). Thus, the Rho IDR governs expression of numerous genes involved in critical functions and required for fitness in the gut.

Discussion

We have uncovered a molecular mechanism employed by *B. thetaiotaomicron*, and potentially other commensals, to successfully colonize the mammalian gut. We determined that the IDR of *B. thetaiotaomicron* Rho (Fig. 1) promotes LLPS (Figs. 2 and 3) and that this promotion, in turn, alters Rho's ability to terminate transcription (Fig. 4), thereby modifying expression of numerous genes when *B. thetaiotaomicron* is present in the gut (Fig. 5).

The IDR increases Rho's termination activity by promoting LLPS but is not required for Rho's core biochemical function (Fig. 4). That LLPS-promoting conditions increase termination by WT Rho may reflect higher activity resulting from an increase in the local Rho concentration and/or an expanded recognition of RNA templates mediated by the IDR. This would enable *B. thetaiotaomicron* to terminate certain RNAs (exemplified by the *mgtA* RNA in our experiments) in all conditions because they harbor sequences readily recognized by Rho (i.e., *rut* sites and secondary binding sites) and other RNAs primarily in LLPS-promoting conditions because they are poorly recognized by Rho (exemplified by the *roc* RNA in our experiments) in non-LLPS conditions (Fig. 6).

Carbon limitation decreases ribosome abundance (*38*), thereby reducing mRNA protection by ribosomes and favoring Rho interaction with unprotected mRNAs. Carbon limitation also elicits Rho LLPS in an IDR-dependent manner (Fig. 3). The IDR is located immediately adjacent to Rho's primary RNA binding site (fig. S1) and interacts with RNAs resulting in LLPS (Fig. 2). We would like to suggest that the IDR enhances Rho's termination activity by increasing the likelihood that Rho will bind an RNA rather than affecting subsequent steps in the transcription termination process (*39*).

Colonization of the mammalian gut is a complex process that includes competition with other bacteria for scarce nutrients and resistance to the host's immune system. The ability of an organism to respond to such a rapidly changing environment ensures its fitness advantage. By triggering sequestration of Rho molecules in an LLPS compartment, carbon starvation and the murine gut enable environmental control of Rho-dependent termination, thereby changing expression of numerous genes. That RNA polymerase and the NusA protein form condensates in *E. coli* (*15*) raises the possibility of additional transcription factors phase separating either independently or in conjunction with Rho in *B. thetaiotaomicron*.

Bacterial evolution is driven by horizontal gene transfer (*40*). Our results highlight how a single, acquired domain within a highly conserved protein expanded its properties without changing its core biochemical function and now plays a critical role in the organism's physiology. Thus, acquisition of the IDR might have allowed *B. thetaiotaomicron* to transition from its original milieu to the mammalian gut. Our findings establish a critical role for LLPS in host-commensal bacteria interactions, expanding LLPS importance for bacterial physiology (*11*–*13*) and defining new targets for gut microbiota manipulation.

REFERENCES AND NOTES

1. J. Durack, S. V. Lynch, *J. Exp. Med.* **216**, 20–40 (2018).
2. P. J. Turnbaugh *et al.*, *Nature* **444**, 1027–1031 (2006).
3. R. Hansen *et al.*, *Clin. Transl. Gastroenterol.* **12**, e00287 (2021).
4. P. R. Banerjee, A. N. Milin, M. M. Moosa, P. L. Onuchic, A. A. Deniz, *Angew. Chem. Int. Ed.* **56**, 11354–11359 (2017).
5. G. L. Dignon, R. B. Best, J. Mittal, *Annu. Rev. Phys. Chem.* **71**, 53–75 (2020).
6. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
7. Y. Shin, C. P. Brangwynne, *Science* **357**, eaaf4382 (2017).
8. H. J. Wiedner, J. Giudice, *Nat. Struct. Mol. Biol.* **28**, 465–473 (2021).
9. D. S. W. Protter, R. Parker, *Trends Cell Biol.* **26**, 668–679 (2016).
10. J. E. Henninger *et al.*, *Cell* **184**, 207–225.e24 (2021).
11. C. A. Azaldegui, A. G. Vecchiarelli, J. S. Biteen, *Biophys. J.* **120**, 1123–1138 (2021).
12. E. A. Abbondanzieri, A. S. Meyer, *Curr. Genet.* **65**, 691–694 (2019).
13. M. C. Cohan, R. V. Pappu, *Trends Biochem. Sci.* **45**, 668–680 (2020).
14. N. Al-Husini, D. T. Tomares, O. Bitar, W. S. Childers, J. M. Schrader, *Mol. Cell* **71**, 1027–1039.e14 (2018).
15. A. M. Ladouceur *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 18540–18549 (2020).
16. A. Grylak-Mielnicka, V. Bidnenko, J. Bardowski, E. Bidnenko, *Microbiology (Reading)* **162**, 433–447 (2016).
17. A. L. Goodman *et al.*, *Cell Host Microbe* **6**, 279–289 (2009).
18. M. A. Kriner, A. Sevostyanova, E. A. Groisman, *Trends Biochem. Sci.* **41**, 690–699 (2016).
19. Z. Hao, V. Svetlov, E. Nudler, *Transcription* **12**, 171–181 (2021).
20. F. D'Heygère, M. Rabhi, M. Boudvillain, *Microbiology (Reading)* **159**, 1423–1436 (2013).
21. A. H. Yuan, A. Hochschild, *Science* **355**, 198–201 (2017).
22. I. Simon, M. Delaleau, A. Schwartz, M. Boudvillain, *J. Mol. Biol.* **433**, 167060 (2021).
23. W. L. Nowatzke, C. M. Burns, J. P. Richardson, *J. Biol. Chem.* **272**, 2207–2211 (1997).
24. D. Szatmári *et al.*, *Sci. Rep.* **10**, 12002 (2020).
25. S. Maharana *et al.*, *Science* **360**, 918–921 (2018).
26. I. Alshareedah *et al.*, *J. Am. Chem. Soc.* **141**, 14593–14602 (2019).
27. Y. Lin *et al.*, *Cell* **167**, 789–802.e12 (2016).
28. S. Kroschwald, S. Maharana, A. Simon, *Matters* (2017), doi: 10.19185/matters.201702000010.
29. W. B. Schofield, M. Zimmermann-Kogadeeva, M. Zimmermann, N. A. Barry, A. L. Goodman, *Cell Host Microbe* **24**, 120–132.e6 (2018).
30. G. E. Townsend 2nd *et al.*, *mBio* **11**, e03221-19 (2020).
31. K. Hollands *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5376–5381 (2012).
32. R. L. Fuchse, R. L. Millette, W. Zillig, G. Walter, *Eur. J. Biochem.* **3**, 183–193 (1967).
33. L. Carrano *et al.*, *Antimicrob. Agents Chemother.* **42**, 571–578 (1998).
34. G. E. Townsend 2nd *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 233–238 (2018).
35. M. S. Ciampi, *Microbiology (Reading)* **152**, 2515–2528 (2006).
36. M. K. Bjursell, E. C. Martens, J. I. Gordon, *J. Biol. Chem.* **281**, 36269–36279 (2006).
37. P. H. Degnan, N. A. Barry, K. C. Mok, M. E. Taga, A. L. Goodman, *Cell Host Microbe* **15**, 47–57 (2014).
38. R. A. Lazzarini, M. Cashel, J. Gallant, *J. Biol. Chem.* **246**, 4381–4385 (1971).
39. P. Mitra, G. Ghosh, M. Hafeezunnisa, R. Sen, *Annu. Rev. Microbiol.* **71**, 687–709 (2017).
40. H. Ochman, J. G. Lawrence, E. A. Groisman, *Nature* **405**, 299–304 (2000).
41. N. M. Koropatkin, E. C. Martens, J. I. Gordon, T. J. Smith, *Structure* **16**, 1105–1115 (2008).
42. E. C. Martens, H. C. Chiang, J. I. Gordon, *Cell Host Microbe* **4**, 447–457 (2008).
43. B. Langmead, S. L. Salzberg, *Nat. Methods* **9**, 357–359 (2012).
44. B. Langmead, C. Wilks, V. Antonescu, R. Charles, *Bioinformatics* **35**, 421–432 (2019).
45. P. Danecek *et al.*, *Gigascience* **10**, giab008 (2021).
46. A. Roberts, H. Pimentel, C. Trapnell, L. Pachter, *Bioinformatics* **27**, 2325–2329 (2011).
47. A. Roberts, C. Trapnell, J. Donaghey, J. L. Rinn, L. Pachter, *Genome Biol.* **12**, R22 (2011).
48. C. Trapnell *et al.*, *Nat. Biotechnol.* **31**, 46–53 (2013).
49. C. Trapnell *et al.*, *Nat. Biotechnol.* **28**, 511–515 (2010).
50. T. Wu *et al.*, *Innovation (Camb.)* **2**, 100141 (2021).
51. G. Yu, L. G. Wang, Y. Han, Q. Y. He, *OMICS* **16**, 284–287 (2012).
52. A. Paintdakhi *et al.*, *Mol. Microbiol.* **99**, 767–777 (2016).
53. D. N. Mastronarde, *J. Struct. Biol.* **152**, 36–51 (2005).
54. S. Q. Zheng *et al.*, *Nat. Methods* **14**, 331–332 (2017).
55. J. R. Kremer, D. N. Mastronarde, J. R. McIntosh, *J. Struct. Biol.* **116**, 71–76 (1996).
56. J. I. Agulleiro, J. J. Fernandez, *J. Struct. Biol.* **189**, 147–152 (2015).
57. K. Hollands, A. Sevostiyanova, E. A. Groisman, *Proc. Natl. Acad. Sci. U.S.A.* **111**, E1999–E2007 (2014).
58. J. Schindelin *et al.*, *Nat. Methods* **9**, 676–682 (2012).
59. J. Guillén-Boixet *et al.*, *Cell* **181**, 346–361.e17 (2020).
60. M. Varadi *et al.*, *Nucleic Acids Res.* **50**, D439–D444 (2021).

61. J. Jumper *et al*., *Nature* **596**, 583–589 (2021).
62. E. Groisman, Raw data for Bacteria require phase separation for fitness in the gut, Version 1, Mendeley (2023), doi: 10.17632/6w9cng5j6n.1.
63. E. Groisman, Raw data for Bacteria require phase separation for fitness in the gut_2, Version 1, Mendeley (2023), doi: 10.17632/bksgrkfmgw.1.

ACKNOWLEDGMENTS

We would like to thank J. Aronson for valuable comments on the manuscript, anonymous reviewers for constructive feedback on the manuscript, N. Barry for assisting with the mouse experiments, R. Washburn and M. Gottesman for providing bicyclomycin, J. Roberts for providing the anti-Rho antibody, A. DeColli for assisting with the protein size exclusion chromatography, K. Hollands and A. Sevostiyanova for supplying purified *Salmonella* proteins, B. Lim for supplying a plasmid harboring the $_{m}$NeonGreen DNA sequence, the Yale Center for Genomic Analysis for RNA-seq, and the Yale West Campus imaging core for assistance with imaging. **Funding:** This work was funded by the following agencies: National Institutes of Health grants GM123798 (to E.A.G.), GM118159 (to A.L.G.), and AI152421 (to J.L.). **Author contributions:** E.A.G., A.K., G.E.T., and A.L.G. designed the research. A.K. designed and performed experiments: mouse gut colonization competition, phase separation, RNAseq, microscopy, Rho biochemical characterization. G.E.T. designed and performed mouse gut colonization competition experiments. A.K., G.E.T., and X.G. constructed bacterial strains used in the study. S.T. and J.L. designed, performed, and analyzed the cryo-ET experiment. A.K., G.E.T., and E.A.G. analyzed the data. N.P. performed the RNA-seq pathway enrichment analysis. A.K. and E.A.G. wrote and edited the manuscript. E.A.G., A.L.G., and J.L. acquired funding. All authors reviewed the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or supplementary materials; relevant raw data have been deposited at Mendeley (*62*, *63*) RNA-seq data are available in Sequence Read Archive (SRA) with accession number PRJNA931765. Any materials and bacterial strains constructed in this work are available upon request from the corresponding author.

Materials and methods and references cited therein can be found in the main article online.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn7229
Figs. S1 to S16
Tables S1 to S3
MDAR Reproducibility Checklist
Movies S1 and S2

Submitted 15 December 2021; resubmitted 21 December 2022
Accepted 10 February 2023
10.1126/science.abn7229

By **Ali Khaledi-Nasab**

Priced out of science

When I set up my first call with Gloria, a student from Kenya who was having trouble getting into a Ph.D. program in the United States, I assumed she might be struggling with English or not have a competitive CV. I heard of her difficulties through my professional network, and as an international graduate of a U.S. Ph.D. program myself I wanted to help. I could not have been more wrong about her problem. Gloria's English was flawless, and she had a perfect score on the Test of English as a Foreign Language (TOEFL). She had earned a master's degree, publishing a paper along the way. The issue was cost. She had already shelled out a substantial sum for the TOEFL, and now the fees for the eight applications she hoped to submit were more than her family's monthly income. She had sought waivers; all but one university had refused.

Her tale evoked memories of my own struggles applying to graduate schools a decade earlier. I grew up poor in Iran and was the first one in my family to even finish middle school. But I was committed to pursuing my education, including an advanced degree in the United States. I paid for the language tests and application fees using money I had earned by tutoring and was thrilled to receive an offer. But I still needed to buy plane tickets, cover my first month's rent in the United States, and pay a hefty fee to bypass Iran's 2 years of compulsory military service and be allowed to leave the country.

I eventually hit on a drastic solution: I could sell one of my kidneys, which is legal in Iran. I placed an advertisement and found a buyer. I was relieved yet terrified of undergoing a serious operation that might leave me needing medical attention for years to come.

Thankfully, an extremely generous friend intervened. Mostafa had always believed in me and supported me. But I tried to hide my financial difficulties from him, and others, because I was ashamed. So, Mostafa had assumed my family had some emergency education fund for me—until one night when he overheard me talking on the phone with the kidney customer. He asked whether the situation was really that bad. I told him my family could barely afford food. I had no other option—it was either sell an organ or give up on my Ph.D.

The next day, I was astonished to see a large sum deposited into my bank account. Mostafa and his family had pooled all their savings and gifted me the money. The windfall covered all my expenses and even left me with an extra $173, which I carefully stretched until I received my first paycheck in the United States. It was a surreal and life-changing moment that I will always cherish and be grateful for.

"Universities still require application fees and costly standardized tests."

Now that I'm relatively settled and financially secure, I am trying to pay forward the generosity I received by helping students like Gloria—because, unfortunately, not much has changed over the past decade. Universities still require application fees and costly standardized tests, disregarding the financial strain these expenses place on students from low-income households. Moreover, the criteria for waiving application fees too often exclude international students—for instance, by requiring proof of U.S. citizenship or permanent residency.

As for Gloria, two schools ultimately waived the application fees, and she ran through her savings and scraped together funds to apply to four more. I worked with her to prepare a strong application package and reached out to my connections for additional advice. We're still waiting to hear whether she is accepted or this hard-fought money is effectively going down the drain.

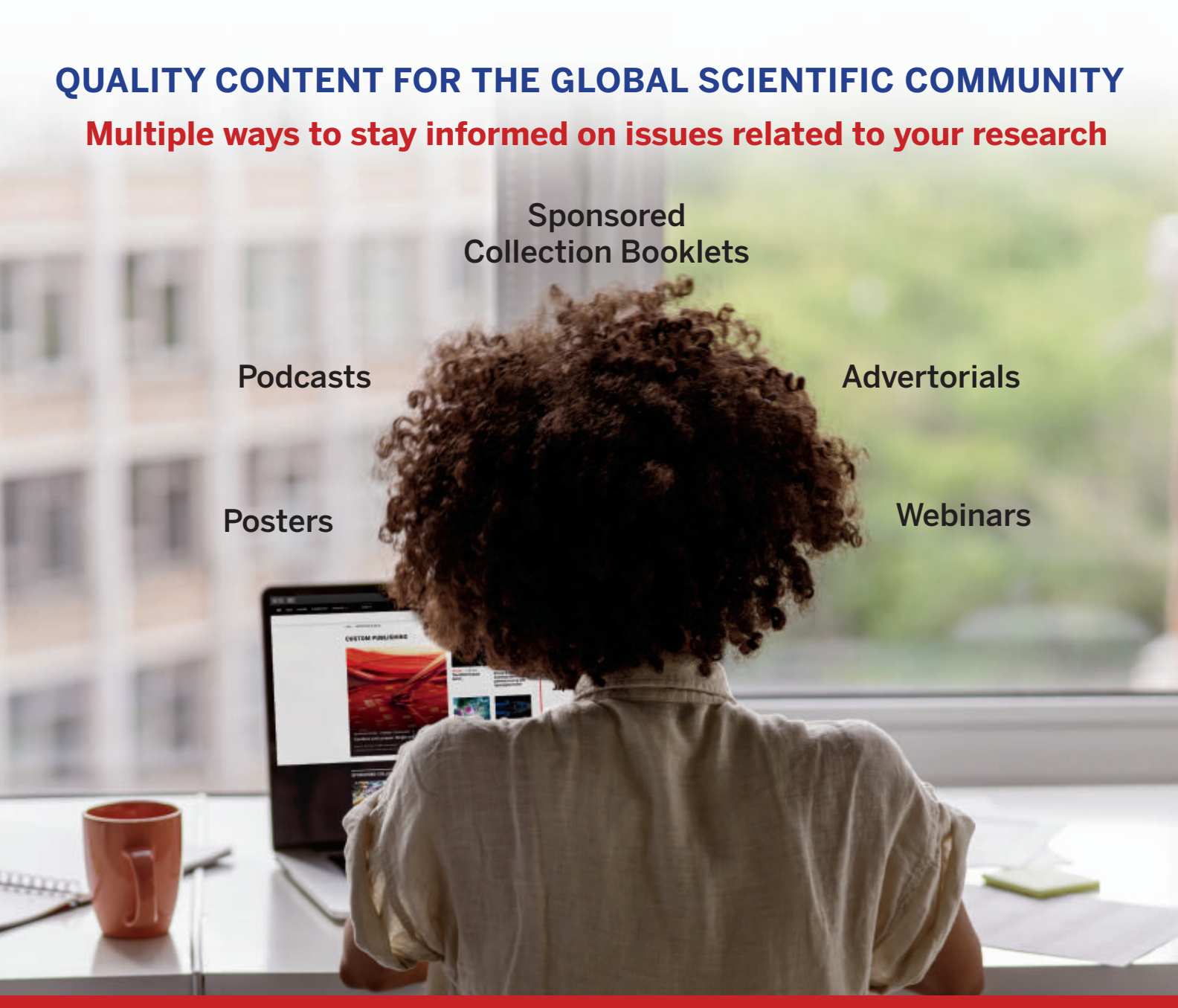
Institutions need to build more equity into the system to provide opportunities for all students, regardless of their background or nationality. They can lower application fees, offer more waivers, or eliminate fees entirely for applicants from economically disadvantaged countries with weak currencies. They can accept more affordable English proficiency tests or cover the costs. They can assist with expenses such as travel and initial living costs. Diversity is not free, and it cannot be left to luck. ■

Ali Khaledi-Nasab is a research scientist at Amazon Web Services.
Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

QUALITY CONTENT FOR THE GLOBAL SCIENTIFIC COMMUNITY

Multiple ways to stay informed on issues related to your research



Science
AAAS

Scan the code and start exploring the latest advances in science and technology innovation!

Science.org/custom-publishing

Brought to you by the *Science*/AAAS Custom Publishing Office.

Posters **Podcasts** **Sponsored Collection Booklets** **Advertorials** **Webinars**

NOMIS & Science
YOUNG EXPLORER AWARD 2023

Research at the intersection of the social and life sciences

Unconventional. Interdisciplinary. Bold.

The **NOMIS & *Science* Young Explorer Award** recognizes and rewards early-career M.D., Ph.D., or M.D./Ph.D. scientists that perform research at the intersection of the social and life sciences. Essays written by these bold researchers on their recent work are judged for clarity, scientific quality, creativity, and demonstration of cross-disciplinary approaches to address fundamental questions.

A cash prize of up to USD 15,000 will be awarded to essay winners, and their engaging essays will be published in ***Science***. Winners will also be invited to share their work and forward-looking perspective with leading scientists in their respective fields at an award ceremony.

Apply by May 15, 2023
at www.science.org/nomis

NOMIS FOUNDATION
Creating the Spark

Science
AAAS