Misconduct suspicions cloud key Alzheimer's research p. 358

Lasing from time crystals pp. 368 & 425

A screening approach to improve cancer nanomedicine pp. 371 & 384

Science

$15
22 JULY 2022
science.org

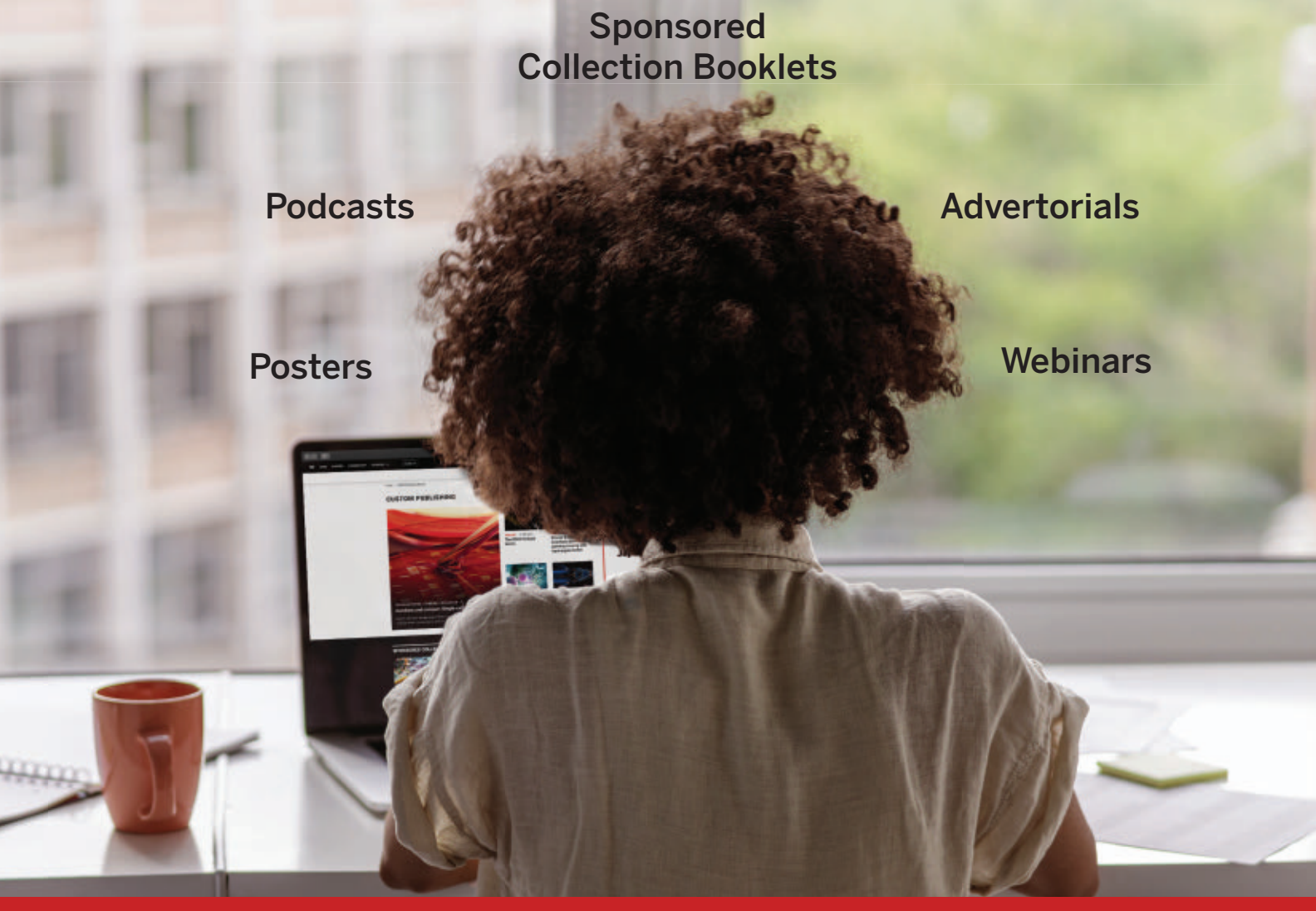
AAAS

SHAPING RANGES

Competition limits bird distributions on tropical mountains p. 416

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**

CONTENTS

22 JULY 2022 • VOLUME 377 • ISSUE 6604

A reference genome for the kākāpō, an endangered parrot in New Zealand, could help conservation breeding programs.

PHOTO: STEPHEN BELCHER/MINDEN PICTURES

NEWS

IN BRIEF

IN DEPTH

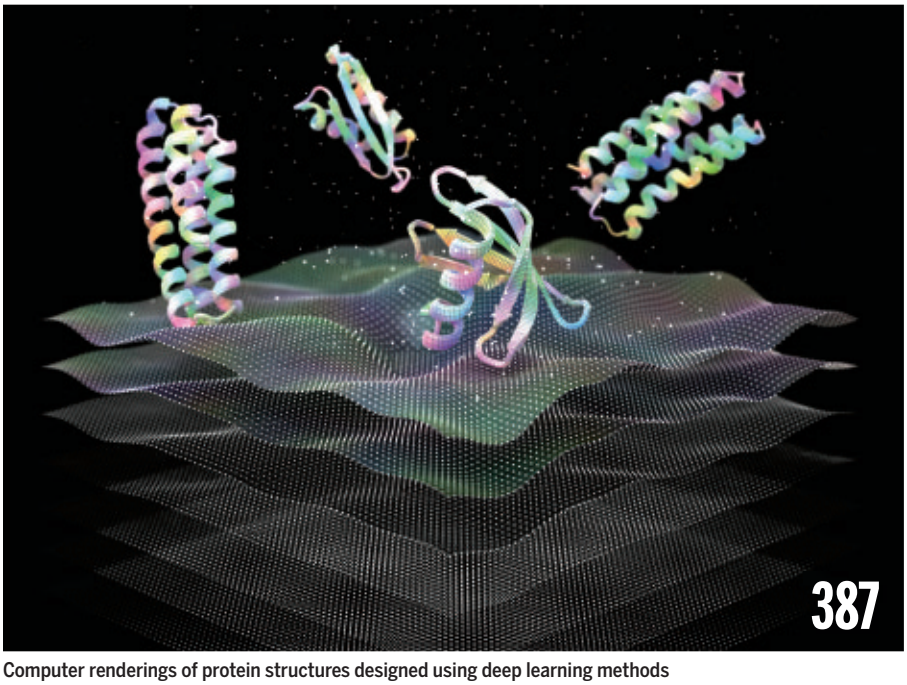
FEATURES

INSIGHTS

PERSPECTIVES

POLICY FORUM

BOOKS *ET AL.*



Computer renderings of protein structures designed using deep learning methods

ON THE COVER

Species tend to live in narrower slices of mountainside on tropical versus temperate mountains. Stronger competition in the tropics explains this pattern for birds. For example, the habitable range of this white-tipped sicklebill (*Eutoxeres aquila*) is limited as a result of competition with its close relative, the buff-tailed sicklebill (*Eutoxeres condamini*). See page 416. *Photo: ©Juan Carlos Vindas/Getty Images*

ILLUSTRATION: IAN HAYDON, INSTITUTE FOR PROTEIN DESIGN

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): $2212; Foreign postage extra: Air assist delivery: $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

EDITORIAL

Confronting 21st-century monkeypox

Michael T. Osterholm is director of the Center for Infectious Disease Research and Policy at the University of Minnesota, Minneapolis, MN, USA. mto@umn.edu

Bruce Gellin is chief of global public health strategy at The Rockefeller Foundation, Washington, DC, USA. bgellin@rockfound.org

The World Health Organization (WHO) hasn't called the current monkeypox outbreak a Public Health Emergency of International Concern (PHEIC), but as a worldwide epidemic, it is clearly an emerging pandemic. More than 12,556 monkeypox cases and three deaths have been reported in 68 countries since early May, and these numbers will rise rapidly with improved surveillance, access to diagnostics, and continuing global spread of infection. Although many tools are needed to control this unfolding pandemic, it's clear that limiting ongoing spread will require a comprehensive international vaccination strategy and adequate supplies.

People 40 years old and younger who have not benefitted from the immunization campaign that eradicated smallpox by 1980 are now susceptible to monkeypox (which is in the same virus family as smallpox), and this lack of population immunity has contributed to the current outbreak. Most of the cases to date have occurred among men who have sex with men (MSM), particularly those with new or multiple partners. Epidemiologic investigations indicate that the predominant mode of transmission is through skin-to-skin and sexual contact, not contact with contaminated clothing or bed linens. Although respiratory droplet transmission might occur, there is no evidence of airborne transmission as there is with COVID-19. And because monkeypox is a self-limited infection with symptoms lasting 2 to 4 weeks, there isn't a chronic carrier state as there is with HIV, which would increase the risk for ongoing transmission.

Although many tools are needed, it is clear that limiting ongoing spread will require widely available vaccination. The ACAM2000 vaccine is licensed by the US Food and Drug Administration for smallpox and allowed for use against monkeypox on an expanded access basis (so-called "compassionate use" for an investigational drug use). It is associated with potentially serious side effects. A newer vaccine with an improved safety profile was approved for monkeypox and smallpox in 2019. This two-dose vaccine, produced by Bavarian Nordic, is a modified vaccinia virus Ankara (MVA; Jynneos in the United States, Imvanex in the European Union, and Imamune in Canada). Its supply, however, is limited.

How can the world leverage these vaccines to control the spread of monkeypox? Transmission among MSM populations must be reduced through aggressive public health measures, including increased vaccination and diagnostic testing and extensive education campaigns targeted at populations at risk and minimizing social stigma. In addition to a massive scaling up of vaccine production, other immediate dose-sparing actions can be taken: administration of a single dose per person instead of two doses (or a first dose followed by a delayed second dose when supplies allow) or intradermal (versus intramuscular) administration of a smaller dose. However, research will be needed to determine whether such dose-sparing approaches provide adequate immune protection.

Determining how vaccine will be allocated to countries and within countries to have the most impact on transmission is essential. Expect major shortages of vaccine among frustrated at-risk individuals for many months to come. To dampen the current outbreak will require vaccination of those at highest risk, with global estimates of the number of MSM ranging from 1 to 3%. The needed global vaccine supply just for MSM is similar to those considered for HIV oral preexposure prophylaxis (PrEP). It is estimated that by 2023, 2.4 million to 5.3 million people worldwide should receive PrEP.

Monkeypox is a zoonotic disease; thus, another critical step is to greatly reduce transmission of the virus from current rodent reservoirs and to prevent spillovers in areas of the world where monkeypox isn't endemic. Long-term control of monkeypox will require vaccinating as many as possible of the 327 million people 40 years of age and younger living in the 11 African countries where monkeypox is endemic in an animal (rodent) reservoir. This effort should include childhood vaccine programs. Surveillance will be needed to identify new animal reservoirs, which might be established in other countries as a result of infected humans inadvertently transmitting the virus to domestic rodents that have subsequent contact with wild rodents.

The smallpox eradication program was a 12-year effort that involved 73 countries working with as many as 150,000 national staff. Because of its animal reservoir, monkeypox can't be eradicated. Unless the world develops and executes an international plan to contain the current outbreak, it will be yet another emerging infectious disease that we will regret not containing.

– Michael T. Osterholm and Bruce Gellin

Published online 19 July 2022; 10.1126/science.add9651

IN BRIEF

Edited by **Jeffrey Brainard**

GLOBAL HEALTH

Pandemic contributes to big drop in childhood vaccinations

In what UNICEF Executive Director Catherine Russell called a "red alert," childhood vaccination rates in many countries worldwide have dropped to the lowest level since 2008, in part because of the COVID-19 pandemic. UNICEF and the World Health Organization together track inoculations against diphtheria, pertussis, and tetanus—which are administered as one vaccine—as a marker for vaccination coverage overall. In 2021, only 81% of children worldwide received the recommended three doses of the combined vaccine, down from 86% in 2019. As a result, some 25 million children remain insufficiently protected against the three dangerous diseases. The majority of children who missed shots live in India, Nigeria, Indonesia, Ethiopia, and the Philippines, but the largest relative drops occurred in two countries with much smaller populations: Myanmar and Mozambique. A similar number of children did not get their first dose of the measles vaccine, and millions also missed polio and human papillomavirus inoculations. The pandemic has limited the ability of health care workers to provide immunizations and disrupted supply chains, UNICEF says; armed conflicts and vaccine misinformation also contributed to the declines.

People queue for vaccinations in June 2021 in India. The country had the largest drop in inoculated children during the pandemic.

PHOTO: SUNIL GHOSH/*HINDUSTAN TIMES*/GETTY IMAGES

At last, U.S. OKs Novavax vaccine

COVID-19 | After a long wait, Novavax's COVID-19 vaccine last week joined the short list of pandemic shots authorized in the United States. The U.S. Food and Drug Administration (FDA) issued an emergency use authorization for the two-dose vaccine on 13 July. Novavax's product is a "protein subunit" vaccine, containing the coronavirus spike protein and an immune stimulant; the company hopes this will appeal to people who worry about side effects from Pfizer's and Moderna's messenger RNA vaccines and Johnson & Johnson's adenovirus-based vaccine. FDA's blessing was delayed in part because Novavax struggled for months to meet the agency's manufacturing standards. The authorization only applies to the primary series of inoculations, but the company hopes that in coming months FDA will authorize a booster dose of the vaccine. Uptake of the Novavax vaccine in the European Union, where it was authorized in December 2021, has been slow: Only about 250,000 people have gotten it so far.

What makes an old growth forest?

CONSERVATION | President Joe Biden's administration last week asked for public comment to help it define what constitutes an "old growth" forest, to inform its efforts to inventory and protect those on federal lands. A better, "universal" definition that reflects evolving scientific understanding of these "unique" ecosystems is needed, says the formal request for comment issued on 15 July by the U.S. Forest Service and the Bureau of Land Management, which together manage nearly 80 million hectares of forests. Although big, ancient trees are often seen as key markers of old growth, forests containing them "differ widely in character" depending on factors such as geography and climate, the agencies note. Suggestions are due by 15 August, and disagreement is likely: Environmentalists want the new definition to be expansive, and the timber industry prefers a narrower one.

U.S. 'superbug' infections rise

COVID-19 | Infections and deaths caused by some of the most harmful antibiotic-resistant pathogens in U.S. hospitals leapt by at least 15% during the first year of the coronavirus pandemic, the Centers for Disease Control and Prevention (CDC) reported last week. The spike boosted 2020 deaths from these infections to 29,400 and was a turnabout from declines in "superbug" infections during the previous decade. Causes included overworked hospital workers forced to let sanitation precautions slip and shortages of personal protective equipment, CDC said. Resistant microbes deemed among the most dangerous drove the largest reported increases in rates of hospital-acquired infections. For example, the rate for *Acinetobacter* bacteria resistant to carbapenem antibiotics increased by 78%, with 7500 such cases. The microbe commonly infects patients on ventilators, such as those hospitalized for COVID-19.

Center tackles ecological data

FUNDING | The University of Colorado, Boulder, will host a new research center to synthesize large amounts of data about environmental change, such as increasing wildfires and biodiversity loss. The Environmental Data Science Innovation and Inclusion Lab will fill "an enormous need," said a statement last week from the U.S. National Science Foundation (NSF) announcing it will fund the new center with $20 million over 5 years. The project will train visiting scientists on computational tools, such as machine learning, to make sense of the vast data being collected by efforts such as the NSF-funded National Ecological Observatory Network and the Ocean Observatories Initiative.

Disputed fossil to return to Brazil

PALEONTOLOGY | A German science ministry plans to repatriate an unusual dinosaur fossil to Brazil, where scientists alleged the artifact had been removed illegally. Scientists at the State Museum of Natural History Karlsruhe (SMNK) published a paper in 2020 describing the chicken-size dinosaur with spearlike feathers that they called *Ubirajara* jubatus. But after the team failed to provide proper documentation of the fossil's legality, the journal withdrew the paper. Last year, a *Science* investigation (1 October 2021, p. 14) prompted the science ministry of Germany's Baden-Württemberg state, which manages SMNK, to investigate; this week, authorities concluded that the museum provided the ministry with false information regarding the fossil's acquisition, prompting the decision to return it.

THEY SAID IT

"I was worried about my goldfish getting too hot. Now I'm worried about the survival of my family and my neighbors."

Hannah Cloke, a natural hazards researcher at the University of Reading, about Europe's worsening heat wave, which she calls "a wake-up call about the climate emergency."

"I want young people to see that this is possible for them, and that it's not off limits because they are Black."

Marine geologist Dawn Wright, in *Nature*, after becoming the first Black person to visit Challenger Deep, Earth's deepest spot, aboard a submersible this month. She and a crewmate used side-scan sonar for seafloor mapping.

Mauna Kea's summit is both a premier site for astronomy and sacred ground.

PHOTO: THOMAS LINKEL/LAIF/REDUX

ASTRONOMY

Native Hawaiians gain voice in managing Mauna Kea

The state of Hawaii this month created a new management body for Mauna Kea, one of the world's best sites for astronomy, that could help resolve a long-running dispute over telescopes on its summit. Many Native Hawaiians consider the mountain sacred and have long objected to the observatories, especially the proposed construction of the Thirty Meter Telescope (TMT), a U.S.-led international project. Under a new state law, control of the summit will be transferred over 5 years from the University of Hawaii to a new body whose 11 members will be appointed by the governor and include representatives of Native Hawaiian groups, the Mauna Kea observatories, and others. Mauna Kea Anaina Hou, one of the Native Hawaiian groups that has opposed TMT's construction, objected that the panel's Native Hawaiian members will not be chosen by the groups and may be heavily outnumbered.

IN DEPTH

A nurse prepares a COVID-19 vaccine in Guwahati, India, on 10 April. A new subvariant named BA.2.75 that was first detected in India has surfaced in many other countries.

PHOTO: ANUPAM NATH/AP IMAGES

COVID-19

As Omicron rages on, virus' path remains unpredictable

Fast-spreading subvariants are coming and going. But an entirely new variant could still emerge

By **Kai Kupferschmidt**

In the short history of the COVID-19 pandemic, 2021 was the year of the new variants. Alpha, Beta, Gamma, and Delta each had a couple of months in the Sun.

But this was the year of Omicron, which swept the globe late in 2021 and has continued to dominate, with subvariants—given more prosaic names such as BA.1, BA.2, and BA.2.12.1—appearing in rapid succession. Two closely related subvariants named BA.4 and BA.5 are now driving infections around the world, but new candidates, including one named BA.2.75, are knocking on the door.

Omicron's lasting dominance has evolutionary biologists wondering what comes next. Some think it's a sign that SARS-CoV-2's initial frenzy of evolution is over and it, like other coronaviruses that have been with humanity much longer, is settling into a pattern of gradual evolution. "I think a good guess is that either BA.2 or BA.5 will spawn additional descendants with more mutations and that one or more of those subvariants will spread and will be the next thing," says Jesse Bloom, an evolutionary biologist at the Fred Hutchinson Cancer Research Center.

But others believe a new variant different enough from Omicron and all other variants to deserve the next Greek letter designation, Pi, may already be developing, perhaps in a chronically infected patient. And even if Omicron is not replaced, its dominance is no cause for complacency, says Maria Van Kerkhove, technical lead for COVID-19 at the World Health Organization. "It's bad enough as it is," she says. "If we can't get people to act [without] a new Greek name, that's a problem."

Even with Omicron, Van Kerkhove emphasizes, the world may face continuing waves of disease as immunity wanes and fresh subvariants arise. She is also alarmed that the surveillance efforts that allowed researchers to spot Omicron and other new variants early on are scaling back or winding down. "Those systems are being dismantled, they are being defunded, people are being fired," she says.

The variants that ruled in 2021 did not arise one out of the other. Instead, they evolved in parallel from SARS-CoV-2 viruses circulating early in the pandemic. In the viral family trees researchers draw to visualize the evolutionary relationships of SARS-CoV-2 viruses, these variants appeared at the tips of long, bare branches. The pattern seems to reflect virus lurking in a single person for a long time and evolving before it emerges and spreads again, much changed.

More and more studies seem to confirm that this occurs in immunocompromised people who can't clear the virus and have long-running infections. On 2 July, for example, Yale University genomic epidemiologist Nathan Grubaugh and his team posted a preprint on medRxiv about one such patient they found accidentally. In the summer of 2021, their surveillance program at the Yale New Haven Hospital kept finding a variant of SARS-CoV-2 called B.1.517 even though that lineage was supposed to have disappeared from the community long ago. All of the samples, it turned out, came from the same person, an immunocompromised patient in his 60s undergoing treatment for a B cell lymphoma. He was infected with B.1.517 in November 2020 and is still positive today.

By following his infection to observe how the virus changed over time, the team found it evolved at twice the normal speed of SARS-CoV-2. (Some of the viruses circulating in the patient today might be qualified as new variants if they were found in the community, Grubaugh says.) That supports the hypothesis that chronic infections could drive the "unpredictable emergence" of new variants, the researchers write in their preprint.

Other viruses that chronically infect patients also change faster within one host than when they spread from one person to the next, says Aris Katzourakis, an evolutionary biologist at the University of Oxford. This is partly a numbers game: There are millions of viruses replicating in an individual, but only a handful are passed on during transmission. So a lot of potential evolution is lost in a chain of infections, whereas a chronic infection allows for endless opportunities to evolve.

But since Omicron emerged in November 2021, no new variants have appeared out

of nowhere. Instead, Omicron has accumulated small changes, making it better at evading immune responses and—together with waning immunity—leading to successive waves. "I think it's probably harder and harder for these new things to emerge and take over because all the different Omicron lineages are stiff competition," Grubaugh says, given how transmissible and immune-evading they already are.

If so, the U.S. decision to update COVID-19 vaccines by adding an Omicron component is the right move, Bloom says; even if Omicron keeps changing, a vaccine based on it is likely to provide more protection than one based on earlier variants.

But it's still possible that an entirely new variant unrelated to Omicron will emerge. Or one of the previous variants, such as Alpha or Delta, could make a comeback after causing a chronic infection and going through a bout of accelerated evolution, says Tom Peacock, a virologist at Imperial College London: "This is what we would call second-generation variants." Given those possibilities, "Studying chronic infections is now more important than ever," says Ravindra Gupta, a microbiologist at the University of Cambridge. "They might tell us the kind of mutational direction the virus will take in the population."

BA.2.75, which was picked up recently, already has some scientists concerned. Nicknamed Centaurus, it evolved from Omicron but seems to have quickly accumulated a whole slew of important changes in its genome, more like an entirely new variant than a new Omicron subvariant. "This looks exactly like Alpha did, or Gamma or Beta," Peacock says.

BA.2.75 appears to be spreading in India, where it was first identified, and has been found in many other countries. Whether it's really outcompeting other subvariants is unclear, Van Kerkhove says: "The data is superlimited right now." "I certainly think it's something worth keeping a close eye on," says Emma Hodcroft, a virologist at the University of Bern.

Keeping an eye on anything is getting harder, however, because surveillance is decreasing. Switzerland, for example, now sequences about 500 samples per week, down from 2000 at its peak, Hodcroft says; the United States went from more than 60,000 per week in January to about 10,000. "Some governments are anxious to cut back on the money they dedicated to sequencing," Hodcroft says. Defending the expense is a "hard sell," she says, "especially if there's a feeling the countries around you will continue sequencing even if you stop."

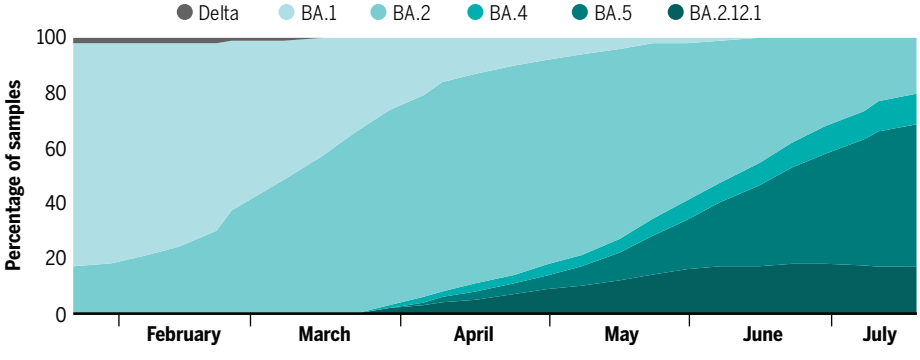
Even if a variant emerges in a place with good surveillance, it may be harder than in the past to predict how big a threat it poses, because differences in past COVID-19 waves, vaccines, and immunization schedules have created a global checkerboard of immunity. That means a new variant might do well in one place but run into a wall of immunity elsewhere. "The situation has become even less predictable," Katzourakis says.

Given that Omicron appears to be milder than previous variants, surveillance efforts should aim to identify variants that cause severe disease in hospitalized patients, Gupta says. "I think that that's where we should be focusing our efforts, because if we keep focusing on new variants genomically, we may get a bit fatigued, and then kind of drop the ball when things do happen."

Many virologists acknowledge that SARS-CoV-2's evolution has caught them by surprise again and again. "It was really in part a failure of imagination," Grubaugh says. But whatever scenario researchers can imagine, Bloom acknowledges the virus will chart its own course: "I think in the end, we just kind of have to wait and see what happens." ■

Making waves

A series of Omicron subvariants has appeared in rapid succession around the world since the beginning of this year. Some scientists say that pattern will likely continue—but an entirely new variant could still arise.



CREDITS: (GRAPHIC) C. BICKEL/*SCIENCE* AND N. DESAI/*SCIENCE*; (DATA) NEXTSTRAIN/GISAID

ATMOSPHERIC SCIENCE

Cleaner air is adding to global warming

Satellites capture fall in light-blocking pollution

By **Paul Voosen**

It's one of the paradoxes of global warming. Burning coal or gasoline releases the greenhouse gases that drive climate change. But it also lofts pollution particles that reflect sunlight and cool the planet, offsetting a fraction of the warming. Now, however, as pollution-control technologies spread, both the noxious clouds and their silver lining are starting to dissipate.

Using an array of satellite observations, researchers have found that the climatic influence of global air pollution has dropped by up to 30% from 2000 levels. Although this is welcome news for public health—airborne fine particles, or aerosols, are believed to kill several million people per year—it is bad news for global warming. The cleaner air has effectively boosted the total warming from carbon dioxide emitted over the same time by anywhere from 15% to 50%, estimates Johannes Quaas, a climate scientist at Leipzig University and lead author of the study. And as air pollution continues to be curbed, he says, "There is a lot more of this to come."

"I believe their conclusions are correct," says James Hansen, a retired NASA climate scientist who first called attention to the "Faustian bargain" of fossil fuel pollution in 1991. He says it's impressive scientific detective work because no satellite could directly measure global aerosols over this whole period. "It's like deducing the properties of unobserved dark matter by looking at its gravitational effects." Hansen expects a flurry of follow-up work, as researchers seek to quantify the boost to warming.

Some aerosols, such as black carbon, or soot, absorb heat. But reflective sulfate and nitrate particles have a cooling effect. For many years, they formed from polluting gases escaping from car tailpipes, ship flues, and power plant smokestacks. Technologies to scrub or eliminate this pollution have spread slowly from North America and Europe to the developing world. Only in 2010

Pollution in China has fallen with the spread of smokestack scrubbers. But the cleanup is boosting warming.

did air pollution in China begin to decline, for example, and international restrictions on sulfur-heavy ship fuel have come just in the past few years.

The new study, submitted as a preprint to *Atmospheric Chemistry and Physics* in April and expected for publication in the next few months, grew directly out of last year's U.N. climate assessment. It included studies showing aerosol declines in North America and Europe but no clear global trends. Quaas and his co-authors thought two NASA satellites, Terra and Aqua, operating since 1999 and 2002, might be able to help.

The satellites tally Earth's incoming and outgoing radiation, which has enabled several research groups, including Quaas and his colleagues, to track the increase in infrared heat trapped by greenhouse gases. But one instrument on Aqua and Terra has also shown a decline in reflected light. Models suggested a decrease in aerosols is partly responsible, says Venkatachalam Ramaswamy, director of the National Oceanic and Atmospheric Administration's Geophysical Fluid Dynamics Laboratory. "It's very hard to find alternate reasons for this," he says.

Quaas and his co-authors have now taken things a step further with two instruments on Terra and Aqua that record the haziness of the sky—and therefore its aerosol load. From 2000 to 2019, haze over North America, Europe, and East Asia clearly declined, although it continued to thicken over coal-dependent India.

Aerosols don't just reflect light on their own; they can also alter clouds. By serving as nuclei on which water vapor condenses, pollution particles reduce cloud droplet size and increase their number, making clouds more reflective. Reducing pollution should undo the effect—and using the same instruments, Quaas and his team found a clear decrease in cloud droplet concentrations in the same regions where aerosols declined.

The evidence in the paper is clear, says Joyce Penner, an atmospheric scientist at the University of Michigan, Ann Arbor. "It's remarkable that we're seeing this already," she says. "This is contributing a lot to the climate changes we're seeing in the current era."

Just how much this declining reflectivity has boosted recent warming is hard to quantify, says Stuart Jenkins, a doctoral student at the University of Oxford who is also studying the aerosol decline. In forthcoming work, Jenkins will show there's just too much natural variability in the past 20 years to pick out the effect of clearer skies.

Whatever the exact contribution, it is sure to grow as air quality continues to improve around the world. The answer isn't to keep polluting, says Jan Cermak, a remote-sensing scientist at the Karlsruhe Institute of Technology. "Air pollution kills people. We need clean air. There is no question about that." Instead, efforts to reduce greenhouse gases need to be redoubled, he says.

But with Earth having warmed by some 1.2°C since preindustrial times, Hansen thinks there's little hope of cutting emissions fast enough to meet the 1.5°C target he and other scientists have called for. And so the solution, he says, could come back to aerosols, this time ones spread deliberately through solar geoengineering—the controversial idea of lofting sulfate particles into the stratosphere and creating a global, reflective haze. "It will be necessary to take temporary corrective measures," he says, "almost surely including temporary purposeful use of aerosols to avoid catastrophic implications." ■

GENOMICS

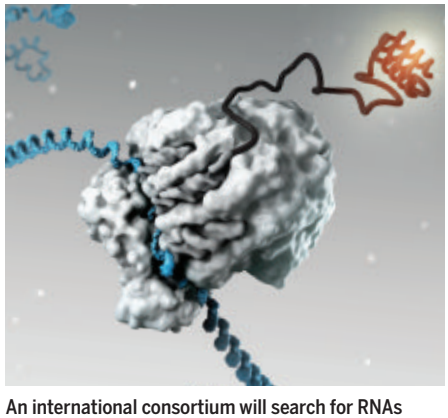
Consortium seeks to expand human gene catalog

Finding sequences that code for short proteins could add thousands of genes

By **Robert F. Service**

The relatively small universe of human genes could grow by up to one-third, if a concerted effort to search for new genes that encode short proteins is successful. Many known miniproteins have already been shown to play key roles in cellular metabolism and disease, so the international effort to catalog new ones and determine their functions, announced last week in *Nature Biotechnology*, could shed light on a vast array of biochemical processes and provide targets for novel medicines.

"The microproteome is a potential gold mine of unexplored biology," says Eric Olson, a molecular biologist at the University of Texas Southwestern Medical Center who is not involved with the new consortium. Anne O'Donnell-Luria, an expert in the genetics of rare diseases at Boston Children's Hospital, adds that the expanded catalog could be a rich source of clues to genetic links to disease. "Everyone will be able to use this data set to make progress in their area."



An international consortium will search for RNAs (blue) that are converted to functional small proteins (orange) by ribosomes (center).

CREDITS: (PHOTO) CHINATOPIX/AP IMAGES; (GRAPHIC) V. ALTOUNIAN/*SCIENCE*

Only 19,370 human genes are known to code for proteins. But current catalogs only include genes for proteins containing at least 100 amino acids each, a cutoff chosen in part because longer DNA sequences make it easier for geneticists to look for commonalities between species. Many smaller proteins are known to exist, but they've largely flown under the radar even though some have been shown to play crucial roles in regulating the immune system, blocking other proteins, and destroying faulty RNAs. "The fact that these have been excluded represents a large hole in genetics and developmental biology," says consortium member John Prensner, a pediatric oncologist at Boston Children's.

When genes are translated into proteins, they are first transcribed into snippets of messenger RNA (mRNA). Cellular organelles called ribosomes then read those mRNA sequences and follow their instructions to string together amino acids into proteins. When scientists scan for genes, they typically look for distinctive DNA sequences flanked by start and stop signals for the protein assembly process, so-called open reading frames (ORFs).

In recent years, researchers have come up with other ways to identify protein-coding sequences. One called Ribo-seq uses high-throughput sequencing technology to catalog all the RNAs in a sample that are bound to a ribosome at a given time. Those RNA sequences point to likely genes, although the technique can't prove that any one sequence makes a stable, functional protein. Ribo-seq databases now contain thousands of ORFs, many of which don't code for known proteins and therefore may represent new ones.

In the consortium's first phase, members scanned seven Ribo-seq databases for candidate ORFs that might correspond with small proteins. After weeding out redundant entries they came up with 7264 candidates. Next, the group will try to identify which of those yield proteins with actual cellular functions. Techniques such as mass spectrometry can help determine whether particular RNAs are translated into stable proteins. Others, such as epitope tagging, use antibodies to track marked proteins, revealing their location and abundance in cells and providing hints about their function.

For now, the 35 investigators involved are funding the effort from their own lab budgets, and don't have immediate plans to seek dedicated funding. "There is so much there, this just needs to be done," says consortium member Sebastiaan van Heesch, a systems biologist at the Princess Máxima Center for Pediatric Oncology in the Netherlands. ■

EUROPE

Russian scientist facing treason charges dies in custody

Advocates say state's zeal for arrests has destroyed the lives of researchers working in sensitive fields

By **Olga Dobrovidova**

Last month, Dmitry Kolker, 54, director of the Laboratory of Quantum Optics at Novosibirsk State University, was dealing with late-stage pancreatic cancer. But on 30 June, agents with Russia's Federal Security Service (FSB) removed him from a cancer clinic, flew him to Moscow, and detained him on charges of treason. By 2 July, he was dead. His family learned of his fate via a curt telegram.

Kolker's colleagues at the Russian Academy of Sciences (RAS) expressed outrage. A group of RAS members signed an open letter protesting FSB's handling of the case and called for "those guilty of our colleague's death to be held accountable." Kolker's family told local media he was accused of leaking state secrets to China. But the RAS group posted a photo of an expert report from an RAS institute concluding that optics lectures Kolker gave in China in 2018 included no classified information.

The case is far from unusual. Three days before Kolker's arrest, FSB arrested another researcher in Siberia: Anatoly Maslov, 75, an aerodynamicist at the Khristianovich Institute of Theoretical and Applied Mechanics, who now faces up to 20 years in prison on treason charges. A 2020 investigation from independent Moscow newspaper *Novaya Gazeta* found that more than 30 scientists had been accused of treason since 2000. Like Maslov, many worked on hypersonics, a research area at the center of a new arms race (*Science*, 10 January 2020, p. 136).

Scientists are "prime targets" for FSB because they have access to sensitive information and often travel to conferences and meet with foreign colleagues, says Ivan Pavlov, a defense lawyer for opposition leader Alexei Navalny's foundation and several treason suspects who fled Russia himself after being detained by FSB. He says the arrests are driven by perverse incentives at FSB, where agents are eager to supply "enemies of the state" in return for bonuses and promotions.

Eugene Chudnovsky, a physicist at Lehman College and co-chair of the Committee of Concerned Scientists, believes the prosecutions may also be "an intimidation tactic" directed at scientists more deeply involved in sensitive research, which the Russian government is careful not to disrupt too much.

Pavlov says the criteria for classifying information as state secrets are purposefully vague, with all details themselves classified, so it is easy to manufacture an accusation. Viktor Kudryavtsev, an aerospace engineer who collaborated with European researchers on a hypersonics project, was arrested in 2018 even though a military review panel had previously approved the work; FSB classified the work 5 years after the project ended.

Laser physicist Dmitry Kolker died this month after being accused of divulging state secrets.

PHOTO: ALEXANDER FEFELOV/REUTERS

The relationship between scientists and the Russian security services has long been fraught, says David Holloway, a historian of the Soviet nuclear program at Stanford University. In the Soviet era, "There was certainly an incentive to find guilty people

and targets to be met," he says. "If you are not arresting people, you aren't doing your job." But during the Cold War, prominent scientists could leverage their usefulness in the nuclear weapons program to gain some protection with party bosses. "The physicists were somehow protected by the bomb, they were needed."

Alexander Fedulov, Kolker's lawyer, says the family intends to fight to clear the physicist's name. Yaroslav Kudryavtsev, a polymer scientist and Viktor's son, also kept up efforts to vindicate his father in court, even after he died in 2021 while under court-mandated travel restrictions. But the family gave up this year, after the Ukraine war began and Russia passed laws to end the jurisdiction of the European Court of Human Rights.

For Chudnovsky, the futility of seeking an acquittal in Russian courts sets these cases apart from the China Initiative in the United States, a law enforcement campaign that was launched in 2018 to prevent China from stealing U.S. technologies and was recently rethought (*Science*, 4 March, p. 945). Still, Pavlov's team has managed to secure pardons and shorter prison terms for several defendants. "In today's Russia, freedom is much more valuable than any available justice," he says.

Private and public support from the scientific community was vital to Viktor Kudryavtsev and his family, but ultimately could not do much to protect the scientist, his son says. Boris Altshuler, a theoretical physicist and human rights activist at RAS's P.N. Lebedev Physical Institute, says that in Soviet times, international pressure from researchers could sometimes bring the security apparatus to heel. "Now, I'm not sure whether the man at the top would listen."

At home, public displays of support have become scarce since the beginning of the war in Ukraine and a government crackdown on protests and dissent. RAS President Alexander Sergeev, who just a few years ago publicly called for Viktor Kudryavtsev to be released from jail, has remained quiet about Kolker and Maslov. In a June speech, he told colleagues to stop "insulting the state" with antiwar declarations.

In Akademgorodok, the enclave of Novosibirsk research institutes where Kolker and Maslov worked, short-lived memorials and graffiti about them keep popping up despite police efforts. At the edge of a forest, a note of protest was taped on top of an official tick warning. It read, "Kolker and Maslov are victims of Moscow occupants, Siberia is not a colony." ■

Olga Dobrovidova is a science journalist in Paris who does climate communications work.

INVASIVE SPECIES

Deadly pest reaches Oregon, sparking fears for ash trees

Emerald ash borer has already killed millions of trees

By **Gabriel Popkin**

Three years ago, forest scientists on the U.S. West Coast launched an effort to gather nearly 1 million seeds of the Oregon ash. The ecologically valuable tree, found from southern California to British Columbia in Canada, often grows along streams and in wetlands, anchoring rich ecosystems.

In part, the collecting effort represented disaster insurance: The emerald ash borer, an invasive, iridescent green beetle that has wiped out ash trees throughout much of the eastern and midwestern United States, was spreading westward, and the saved seeds might one day help restore the species if the pest ever arrived.

The emerald ash borer now has a foothold in the western United States.

Now, it appears that rescue mission began none too soon. Last week, the U.S. Department of Agriculture confirmed that the emerald ash borer (*Agrilus planipennis*) had reached Oregon—and likely has been there for up to 5 years. The discovery marked the borer's first appearance west of the Rocky Mountains; previously it had only gotten as far as Boulder, Colorado. Forest managers now fear for the future of the Oregon ash and at least eight other ash species found only in western North America.

"It's extremely grave and sobering to have the situation upon us," Karen Ripley, a forest health monitoring coordinator with the U.S. Forest Service (USFS), wrote last week in an email to colleagues.

On 30 June, a biologist with the city of Portland alerted officials to adult beetles he saw emerging from a tree in nearby Forest Grove, Oregon. The next day, Wyatt Williams, an invasive species specialist with the Oregon Department of Forestry, confirmed that an Oregon ash was infested. "My heart just sank," he says.

The report opened a new front in the nearly 2-decade-old fight against the borer, an Asian species that was first found in 2002 outside Detroit and has since been documented in 36 states; Washington, D.C.; and parts of Canada.

Once the borer shows up, "You cannot, generally speaking, get rid of [it]," says Leigh Greenwood, a forest specialist at the Nature Conservancy. In Oregon, officials will likely try to slow its spread and reduce its population through selective use of insecticides and by releasing tiny wasps that parasitize and kill the beetles. Such strategies have been used elsewhere with limited success.

In the longer term, some researchers hope to breed trees that can resist the beetle. Of the now-imperiled western ash species, Oregon ash (*Fraxinus latifolia*) is the most immediate concern. In sensitive wetlands where it can form nearly pure stands, no other tree can readily take its place. "In some areas it's the only [tree] species there," says Richard Sniezko, a USFS geneticist and a leader of the seed collecting project.

Sniezko began working on ash trees after attending a 2019 conference. He is now growing seedlings from a number of Oregon ash populations at a research station in the state; colleagues are overseeing a similar set of plantings in Washington and Ohio. Once the ash borer arrives, researchers will observe how the trees fare. Individual trees that hold up better than others might ultimately help scientists breed new, hardier varieties. Such breeding efforts are already underway for other ash species in Ohio (*Science*, 13 November 2020, p. 756).

Researchers are also beginning to collect seeds from up to eight other endemic ash species that only live in the southwestern United States. That effort is challenging because several of the species are rare and grow in remote areas, says Tim Thibault, a curator at the Huntington, a botanical garden in San Marino, California, who is co-leading the project.

The seed collections are only the beginning of a long and expensive process, scientists warn. Rescuing a tree through breeding, Sniezko says, "is not for the faint of heart." ■

PHOTO: MACROSCOPIC SOLUTIONS/SCIENCE SOURCE

The insurrection at the U.S. Capitol on 6 January 2021 showed how politics could motivate violence.

PHOTO: SHAY HORSE/NURPHOTO/GETTY IMAGES

GUN VIOLENCE

Half of Americans anticipate a U.S. civil war soon, survey finds

Findings suggest rising gun violence will spill into the political sphere, driven by conspiracy theories

By **Rodrigo Pérez Ortega**

Violence can seem to be everywhere in the United States, and political violence is in the spotlight, with the 6 January 2021 insurrection as exhibit A. Now, a large study confirms one in five Americans believes violence motivated by political reasons is—at least sometimes—justified. Nearly half expect a civil war, and many say they would trade democracy for a strong leader, a preprint submitted last week to medRxiv found.

"This is not a study that's meant to shock," says Rachel Kleinfeld, a political violence expert at the Carnegie Endowment for International Peace who was not involved in the research. "But it should be shocking."

Firearm deaths in the United States grew by nearly 43% between 2010 and 2020, and gun sales surged during the coronavirus pandemic. Garen Wintemute, an emergency medicine physician and longtime gun violence researcher at the University of California, Davis, wondered what those trends portend for civil unrest. "Sometimes being an ER [emergency room] doc is like being the bow man on the *Titanic* going, 'Look at that iceberg!'" he says.

He and his colleagues surveyed more than 8600 adults in English and Spanish about their views on democracy in the United States, racial attitudes in U.S. society, and their own attitudes toward political violence. The respondents were part of the Ipsos KnowledgePanel—an online research panel that has been used widely, including by Wintemute for research on violence and firearm ownership. The team then applied statistical methods to extrapolate the survey results to the entire country.

Although almost all respondents thought it's important for the United States to remain a democracy, about 40% said having a strong leader is more important. Half expect a civil war in the United States in the next few years. (The survey didn't specify when.) "The fact that basically half the country is expecting a civil war is just chilling," Wintemute says. And many expect to take part. If found in a situation where they think violence is justified to advance an important political objective, about one in five respondents thinks they will likely be armed with a gun. About 7% of participants—which would correspond to about 18 million U.S. adults—said they would be willing to kill a person in such a situation.

Kleinfeld says the study's findings are compelling because of the large number of participants and because it asked about specific scenarios in which participants think violence is justified—such as for self-defense or to stop people with different political beliefs from voting. The sample does slightly overrepresent older people, who are not known to commit much violence worldwide, she says. "So the fact that you're [still] getting these high numbers … is really quite concerning."

She is less alarmed by the shaky support for democracy, noting that political gridlock—as in U.S. politics today—can often distort attitudes. "What people mean by 'democracy' is pretty fuzzy," she says. Political paralysis, she adds, can quickly lead people who think, "Yeah, I like democracy," to also say, "Yeah, I want a strong man" in leadership.

"The findings are scary, but not surprising," Kurt Braddock, who studies the psychology of extremist communication at American University, wrote in an email to *Science*. In recent years, he says, the United States has seen an increase in individual willingness to engage in violence—homicides in cities increased 44% between 2019 and 2021, for instance—an attitude he says is likely to spill into the political sphere.

Researchers have criticized the sampling and survey methodology of previous studies that found increasing support for political violence. But the new study generally agrees with earlier efforts, Kleinfeld says. A small survey from 2021, for instance, found about 46% of voters thought the United States would have another civil war, and another showed more than one-third of Americans agree that "The traditional American way of life is disappearing so fast that we may have to use force to save it."

Wintemute and colleagues found that conspiracy theories, some rooted in racism, are helping shape views about political violence. They found roughly two in five adults agreed with the white nationalist "great replacement theory," or the idea that native-born white voters are being replaced by immigrants for electoral gains. And one in five respondents believed the false QAnon conspiracy theory that U.S. institutions are controlled by an elite group of Satan-worshipping pedophiles.

To reduce the threat of political violence, Braddock says, the first step is to call out the disinformation online and in right-wing media, some of which is taken directly from extremist propaganda. "We need to call that out for what it is before we can begin to address the problems it is causing." Kleinfeld adds that leaders—from politicians and media personalities to church pastors—can also make a difference. Experiments show courageous leaders can deter their communities from engaging in violence. "Now's the time to take this seriously and not put our heads in the sand," Kleinfeld says. ■

BLOTS ON A FIELD?

A neuroscience image sleuth finds signs of fabrication in scores of Alzheimer's articles, threatening a reigning theory of the disease

By **Charles Piller**

In August 2021, Matthew Schrag, a neuroscientist and physician at Vanderbilt University, got a call that would plunge him into a maelstrom of possible scientific misconduct. A colleague wanted to connect him with an attorney investigating an experimental drug for Alzheimer's disease called Simufilam. The drug's developer, Cassava Sciences, claimed it improved cognition, partly by repairing a protein that can block sticky brain deposits of the protein amyloid beta (Aβ), a hallmark of Alzheimer's. The attorney's clients—two prominent neuroscientists who are also short sellers who profit if the company's stock falls—believed some research related to Simufilam may have been "fraudulent," according to a petition later filed on their behalf with the U.S. Food and Drug Administration (FDA).

Schrag, 37, a softspoken, nonchalantly rumpled junior professor, had already gained some notoriety by publicly criticizing the controversial FDA approval of the anti-Aβ drug Aduhelm. His own research also contradicted some of Cassava's claims. He feared volunteers in ongoing Simufilam trials faced risks of side effects with no chance of benefit.

Neuroscientist and physician Matthew Schrag found suspect images in dozens of papers involving Alzheimer's disease, including Western blots (projected in green) measuring a protein linked to cognitive decline in rats.

So he applied his technical and medical knowledge to interrogate published images about the drug and its underlying science—for which the attorney paid him $18,000. He identified apparently altered or duplicated images in dozens of journal articles. The attorney reported many of the discoveries in the FDA petition, and Schrag sent all of them

PHOTO: JOSEPH ROSS

to the National Institutes of Health (NIH), which had invested tens of millions of dollars in the work. (Cassava denies any misconduct [see sidebar, p. *363*].)

But Schrag's sleuthing drew him into a different episode of possible misconduct, leading to findings that threaten one of the most cited Alzheimer's studies of this century and numerous related experiments.

The first author of that influential study, published in *Nature* in 2006, was an ascending neuroscientist: Sylvain Lesné of the University of Minnesota (UMN), Twin Cities. His work underpins a key element of the dominant yet controversial amyloid hypothesis of Alzheimer's, which holds that Aβ clumps, known as plaques, in brain tissue are a primary cause of the devastating illness, which afflicts tens of millions globally. In what looked like a smoking gun for the theory and a lead to possible therapies, Lesné and his colleagues discovered an Aβ subtype and seemed to prove it caused dementia in rats. If Schrag's doubts are correct, Lesné's findings were an elaborate mirage.

Schrag, who had not publicly revealed his role as a whistleblower until this article, avoids the word "fraud" in his critiques of Lesné's work and the Cassava-related studies and does not claim to have proved misconduct. That would require access to original, complete, unpublished images and in some cases raw numerical data. "I focus on what we can see in the published images, and describe them as red flags, not final conclusions," he says. "The data should speak for itself."

A 6-month investigation by *Science* provided strong support for Schrag's suspicions and raised questions about Lesné's research. A leading independent image analyst and several top Alzheimer's researchers—including George Perry of the University of Texas, San Antonio, and John Forsayeth of the University of California, San Francisco (UCSF)—reviewed most of Schrag's findings at *Science*'s request. They concurred with his overall conclusions, which cast doubt on hundreds of images, including more than 70 in Lesné's papers. Some look like "shockingly blatant" examples of image tampering, says Donna Wilcock, an Alzheimer's expert at the University of Kentucky.

The authors "appeared to have composed figures by piecing together parts of photos from different experiments," says Elisabeth Bik, a molecular biologist and well-known forensic image consultant. "The obtained experimental results might not have been the desired results, and that data might have been changed to … better fit a hypothesis."

Early this year, Schrag raised his doubts with NIH and journals including *Nature*; two, including *Nature* last week, have published expressions of concern about papers by Lesné. Schrag's work, done independently of Vanderbilt and its medical center, implies millions of federal dollars may have been misspent on the research—and much more on related efforts. Some Alzheimer's experts now suspect Lesné's studies have misdirected Alzheimer's research for 16 years.

"The immediate, obvious damage is wasted NIH funding and wasted thinking in the field because people are using these results as a starting point for their own experiments," says Stanford University neuroscientist Thomas Südhof, a Nobel laureate and expert on Alzheimer's and related conditions.

Lesné did not respond to requests for comment. A UMN spokesperson says the university is reviewing complaints about his work.

"You can't cheat to cure a disease. Biology doesn't care."

Matthew Schrag, Vanderbilt University

To Schrag, the two disputed threads of Aβ research raise far-reaching questions about scientific integrity in the struggle to understand and cure Alzheimer's. Some adherents of the amyloid hypothesis are too uncritical of work that seems to support it, he says. "Even if misconduct is rare, false ideas inserted into key nodes in our body of scientific knowledge can warp our understanding."

IN HIS MODEST OFFICE, steps away from a buzzing refrigerator, Schrag displays an antique microscope—an homage to predecessors who applied painstaking bench science to medicine's endless enigmas. A small sign on his desk reads, "Everything is figureoutable."

So far, Alzheimer's has been an exception. But Schrag's background has left him comfortable with the field's contradictions. His father hails from a family of Mennonites, known for their philosophy of peacemaking—but joined the military. The family moved from Arizona to Germany to England before settling in Davenport, a tiny cow town in eastern Washington. After leaving the Air Force, Schrag's dad became a nurse and worked in a nursing home. As a young teen, Schrag volunteered to visit dementia patients there. "I remembered being mystified by a lot of the strange behaviors," he says. It was a formative experience "to see people struggling with such unfair symptoms."

Home-schooled by his mom, Schrag entered community college at 16, like many of the town's studious kids—including his teenage sweetheart and future wife, Sarah. They now live on a small ranch outside Nashville with their two young children and three aging horses that Sarah grew up with.

While prepping for medical school at the University of North Dakota, Schrag spent long hours in a neuropharmacology lab absorbing the patient rhythms of science. He repeated experiments over and over, refining his skills. These included a protein identification method known as the Western blot. It uses electricity to drive protein-rich tissue samples through a gel that acts like a sieve to separate the molecules by size. Distinct proteins, tagged and illuminated by fluorescent antibodies, appear as stacked bands.

In 2006, Schrag's first publication examined how feeding a high-cholesterol diet to rabbits seemed to increase Aβ plaques and iron deposits in one part of their brains. Not long afterward, when he was an M.D.-Ph.D. student at Loma Linda University, another research group found support for a link between Alzheimer's and iron metabolism. Encouraged, Schrag poured his energy into trying to confirm the connection in people—and failed. The experience introduced him to a disquieting element of Alzheimer's research. With this enigmatic, complex disease, even careful experiments done in good faith can fail to replicate, leading to dead ends and unexpected setbacks.

One of its biggest mysteries is also its most distinctive feature: the plaques and other protein deposits that German pathologist Alois Alzheimer first saw in 1906 in the brain of a deceased dementia patient. In 1984, Aβ was identified as the main component of the plaques. And in 1991, researchers traced family-linked Alzheimer's to mutations in the gene for a precursor protein from which amyloid derives. To many scientists, it seemed clear that Aβ buildup sets off a cascade of damage and dysfunction in neurons, causing dementia. Stopping amyloid deposits became the most plausible therapeutic strategy.

Hundreds of clinical trials of amyloid-targeted therapies have yielded few glimmers of promise, however; only the underwhelming Aduhelm has gained FDA approval. Yet Aβ still dominates research and drug development. NIH spent about $1.6 billion on projects that mention amyloids in this fiscal year, about half its overall Alzheimer's funding. Scientists who advance other potential Alzheimer's causes, such as immune dysfunction or inflammation, complain they have been sidelined by the "amyloid mafia." Forsayeth says the amyloid hypothesis became "the scientific equivalent of the Ptolemaic model of the Solar System," in which the Sun and planets rotate around Earth.

By 2006, the centenary of Alois Alzheimer's epic discovery, a growing cadre of skeptics wondered aloud whether the field needed a reset. Then, a breathtaking *Nature* paper entered the breach.

It emerged from the lab of UMN physi-

cian and neuroscientist Karen Ashe, who had already made a remarkable series of discoveries. As a medical resident at UCSF, she contributed to Nobel laureate Stanley Prusiner's pioneering work on prions—infectious proteins that cause rare neurological disorders. In the mid-1990s, she created a transgenic mouse that churns out human Aβ, which forms plaques in the animal's brain. The mouse also shows dementia-like symptoms. It became a favored Alzheimer's model.

By the early 2000s, "toxic oligomers," subtypes of Aβ that dissolve in some bodily fluids, had gained currency as a likely chief culprit for Alzheimer's—potentially more pathogenic than the insoluble plaques. Amyloid oligomers had been linked to impaired communication between neurons in vitro and in animals, and autopsies have shown higher levels of the oligomers in people with Alzheimer's than in cognitively sound individuals. But no one had proved that any one of the many known oligomers directly caused cognitive decline.

Sylvain Lesné, University of Minnesota, Twin Cities

In the brains of Ashe's transgenic mice, the UMN team discovered a previously unknown oligomer species, dubbed Aβ*56 (pronounced "amyloid beta star 56") after its relatively heavy molecular weight compared with other oligomers. The group isolated Aβ*56 and injected it into young rats. The rats' capacity to recall simple, previously learned information—such as the location of a hidden platform in a maze—plummeted. The 2006 paper's first author, sometimes credited as the discoverer of Aβ*56, was Lesné, a young scientist Ashe had hired straight out of a Ph.D. program at the University of Caen Normandy in France.

Ashe touted Aβ*56 on her website as "the first substance ever identified in brain tissue in Alzheimer's research that has been shown to cause memory impairment." An accompanying editorial in *Nature* called Aβ*56 "a star suspect" in Alzheimer's. Alzforum, a widely read online hub for the field, titled its coverage, "Aβ Star is Born?" Less than 2 weeks after the paper was published, Ashe won the prestigious Potamkin Prize for neuroscience, partly for work leading to Aβ*56.

The *Nature* paper has been cited in about 2300 scholarly articles—more than all but four other Alzheimer's basic research reports published since 2006, according to the Web of Science database. Since then, annual NIH support for studies labeled "amyloid, oligomer, and Alzheimer's" has risen from near zero to $287 million in 2021. Lesné and Ashe helped spark that explosion, experts say.

The paper provided an "important boost" to the amyloid and toxic oligomer hypotheses when they faced rising doubts, Südhof says. "Proponents loved it, because it seemed to be an independent validation of what they have been proposing for a long time."

"That was a really big finding that kind of turned the field on its head," partly because of Ashe's impeccable imprimatur, Wilcock says. "It drove a lot of other investigators to … go looking for these [heavier] oligomer species."

As Ashe's star burned more brightly, Lesné's rose. He joined UMN with his own NIH-funded lab in 2009. Aβ*56 remained a primary research focus. Megan Larson, who worked as a junior scientist for Lesné and is now a product manager at Bio-Techne, a biosciences supply company, calls him passionate, hardworking, and charismatic. She and others in the lab often ran experiments and produced Western blots, Larson says, but in their papers together, Lesné prepared all the images for publication.

He became a leader of UMN's neuroscience graduate program in 2020, and in May 2021, 4 months after Schrag delivered his concerns to NIH, Lesné received a coveted R01 grant from the agency, with up to 5 years of support. The NIH program officer for the grant, Austin Yang—a co-author on the 2006 *Nature* paper—declined to comment.

IN DECEMBER 2021, Schrag visited PubPeer, a website where scientists flag possible errors in published papers. Many of the site's posts come from technical gumshoes who deconstruct Western blots for telltale marks indicating that bands representing proteins could have been removed or inserted where they don't belong. Such manipulations can falsely suggest a protein is present—or alter the levels at which a detected protein is apparently found. Schrag, still focused on Cassava-linked scientists, was looking for examples that could refine his own sleuthing.

In a PubPeer search for "Alzheimer's," postings about articles in *The Journal of Neuroscience* caught Schrag's eye. They questioned the authenticity of blots used to differentiate Aβ and similar proteins in mouse brain tissue. Several bands seemed to be duplicated. Using software tools, Schrag confirmed the PubPeer comments and found similar problems with other blots in the same articles. He also found some blot backgrounds that seemed to have been improperly duplicated.

Three of the papers listed Lesné, whom Schrag had never heard of, as first or senior author. Schrag quickly found that another Lesné paper had also drawn scrutiny on PubPeer, and he broadened his search to Lesné papers that had not been flagged there. The investigation "developed organically," he says, as other apparent problems emerged.

"So much in our field is not reproducible, so it's a huge advantage to understand when data streams might not be reliable," Schrag says. "Some of that's going to happen reproducing data on the bench. But if it can happen in simpler, faster ways—such as image analysis—it should." Eventually Schrag ran across the seminal *Nature* paper, the basis for many others. It, too, seemed to contain multiple doctored images.

Science asked two independent image analysts—Bik and Jana Christopher—to review Schrag's findings about that paper and others by Lesné. They say some supposed manipulation might be digital artifacts that can occur inadvertently during image processing, a possibility Schrag concedes. But Bik found his conclusions compelling and sound. Christopher concurred about the many duplicated images and some markings suggesting cut-and-pasted Western blots flagged by Schrag. She also identified additional dubious blots and backgrounds he had missed.

In the 16 years following the landmark paper, Lesné and Ashe—separately or jointly—published many articles on their stellar oligomer. Yet only a handful of other groups have reported detecting Aβ*56.

Citing the ongoing UMN review of Lesné's work, Ashe declined via email to be interviewed or to answer written questions posed by *Science*, which she called "sobering." But she wrote, "I still have faith in Aβ*56," noting her ongoing work studying the structure of Aβ oligomers. "We have promising initial results. I remain excited about this work, and believe it has the potential to explain why Aβ therapies may yet work despite recent failures targeting amyloid plaques."

But even before Schrag's investigation, the spotty evidence that Aβ*56 plays a role in Alzheimer's had raised eyebrows. Wilcock has long doubted studies that claim to use "purified" Aβ*56. Such oligomers are notoriously unstable, converting to other oligomer types spontaneously. Multiple types can be present in a sample even after purification efforts, making it hard to say any cognitive effects are due to AB*56 alone, she notes—assuming it exists. In fact, Wilcock and others say, several labs have tried and failed to find Aβ*56, although few have published those findings. Journals are often uninterested in negative

ILLUSTRATION: N. DESAI/*SCIENCE*

results, and researchers can be reluctant to contradict a famous investigator.

An exception was Harvard University's Dennis Selkoe, a leading advocate of the amyloid and toxic oligomer hypotheses, who has cited the *Nature* paper at least 13 times. In two 2008 papers, Selkoe said he could not find Aβ*56 in human fluids or tissues.

Selkoe examined Schrag's dossier on Lesné's papers at *Science*'s request, and says he finds it credible and well supported. He did not see manipulation in every suspect image, but says, "There are certainly at least 12 or 15 images where I would agree that there is no other explanation" than manipulation. One—an image in the *Nature* paper displaying purified Aβ*56—shows "very worrisome" signs of tampering, Selkoe says. The same image reappeared in a different paper, co-authored by Lesné and Ashe, 5 years later. Many other images in Lesné's papers might be improper—more than enough to challenge the body of work, Selkoe adds.

A few of Lesné's questioned papers describe a technique he developed to measure Aβ oligomers separately in brain cells, spaces outside the cells, and cell membranes. Selkoe recalls Ashe talking about her "brilliant postdoctoral fellow" who devised it. He was skeptical of Lesné's claim that oligomers could be analyzed separately inside and outside cells in a mixture of soluble material from frozen or processed brain tissue. "All of us who heard about that knew in a moment that it made no biochemical sense. If it did, we'd all be using a method like that," Selkoe says. The *Nature* paper depended on that method.

Selkoe himself co-authored a 2006 paper with Lesné in the *Annals of Neurology*. They sought to neutralize the effects of toxic oligomers, although not Aβ*56. The paper includes an image that Schrag, Bik, and Christopher agree was reprinted as if original in two subsequent Lesné articles. Selkoe calls that "highly egregious."

Given those findings, the scarcity of independent confirmation of the Aβ*56 claims seems telling, Selkoe says. "In science, once you publish your data, if it's not readily replicated, then there is real concern that it's not correct or true. There's precious little clear-cut evidence that Aβ*56 exists, or if it exists, correlates in a reproducible fashion with features of Alzheimer's—even in animal models."

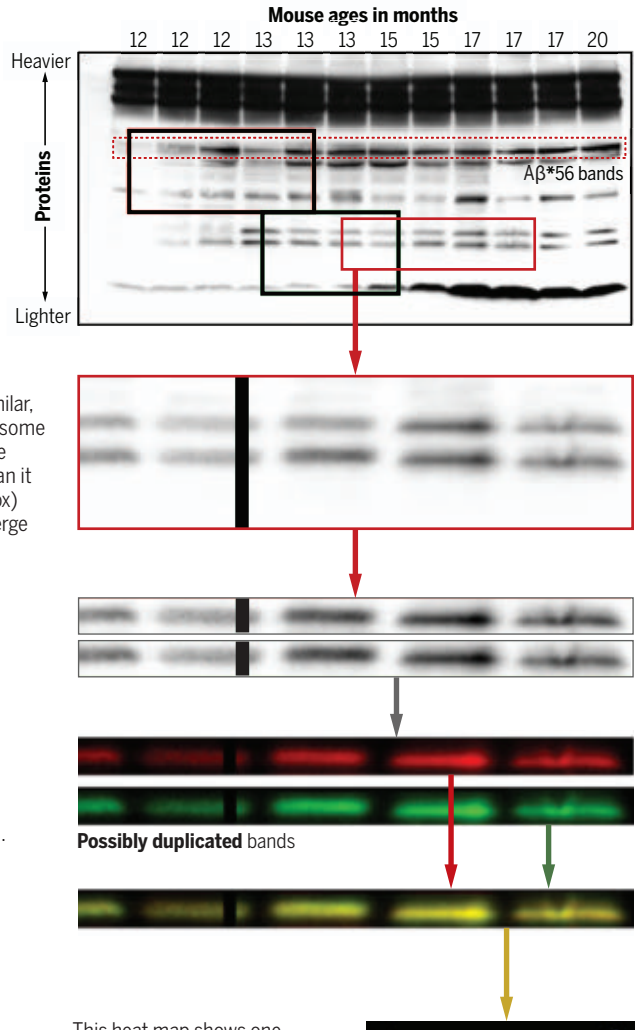
IN ALL, SCHRAG OR BIK identified more than 20 suspect Lesné papers; 10 concerned Aβ*56. Schrag contacted several of the journals starting early this year, and Lesné and his collaborators recently published two corrections. One for a 2012 paper in *The Journal of Neuroscience* replaced several images Schrag had flagged as problematic, writing that the earlier versions had been

How an image sleuth uncovered possible tampering

Vanderbilt University neuroscientist Matthew Schrag found apparently falsified images in papers by University of Minnesota, Twin Cities, neuroscientist Sylvain Lesné, including a 2006 paper in *Nature* co-authored with Karen Ashe and others. It linked an amyloid-beta (Aβ) protein, Aβ*56, to Alzheimer's dementia.

Image in question

Ashe uploaded this Western blot to PubPeer after Schrag said the version published in *Nature* showed cut marks suggesting improper tampering with bands portraying Aβ*56 and other proteins (black boxes added by Ashe). The figure shows levels of Aβ*56 (dashed red box) increasing in older mice as symptoms emerge. But Schrag's analysis suggests this version of the image contains improperly duplicated bands.



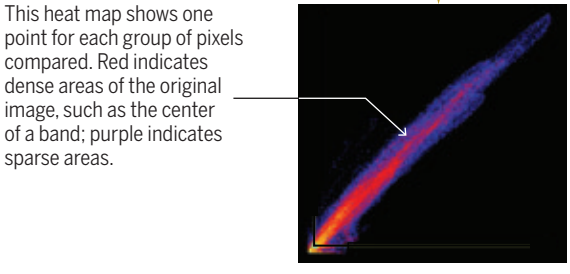
1 Spot the similarities
Some bands looked abnormally similar, an apparent manipulation that in some cases (not shown) could have made Aβ*56 appear more abundant than it was. One striking example (red box) ostensibly shows proteins that emerge later in the life span than Aβ*56.

2 Match contrast
Schrag matched the contrast level in the two sets of bands for an apples-to-apples comparison.

3 Colorize and align
Schrag turned backgrounds black to make the bands easier to see, then colorized them and precisely matched their size and orientation.

4 Merge
He merged the sets of colorized bands. The areas of the image that are identical appear in yellow.

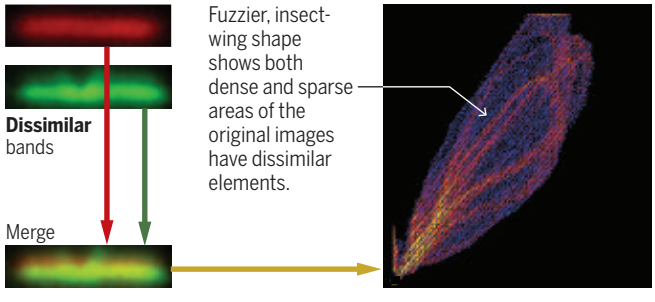
5 Calculate similarity
Schrag then calculated the correlation coefficient, showing the strength of the relationship between the merged bands. Identical images show a correlation of 1, and display as a straight 45° angle line. These bands show a 0.98 correlation, highly unlikely to occur by chance.



This heat map shows one point for each group of pixels compared. Red indicates dense areas of the original image, such as the center of a band; purple indicates sparse areas.

Unmistakable differences

These images examine dissimilar bands using the same process. In the merged image, clear differences appear in green or red—as expected when comparing naturally produced bands. A degree of correlation is expected, but far lower than in duplicated bands.



Fuzzier, insect-wing shape shows both dense and sparse areas of the original images have dissimilar elements.

CREDITS: (GRAPHIC) C. BICKEL/*SCIENCE*; (DATA) S. LESNÉ *ET AL*., *NATURE* **440**, 352 (2006). HTTPS://DOI.ORG/10.1038/NATURE04533

"processed inappropriately." But Schrag says even the corrected images show numerous signs of improper changes in bands, and in one case, complete replacement of a blot.

A 2013 *Brain* paper in which Schrag had flagged multiple images was also extensively corrected in May. Lesné and Ashe were the first and senior authors, respectively, of the study, which showed "negligible" levels of Aβ*56 in children and young adults, more when people reached their 40s, and steadily increasing levels after that. It concluded that Aβ*56 "may play a pathogenic role very early in the pathogenesis of Alzheimer's disease." The authors said the correction had no bearing on the study's findings.

Schrag isn't convinced. Among other problems, one corrected blot shows multiple bands that appear to have been added or removed artificially, he says.

Karen Ashe, University of Minnesota, Twin Cities

Selkoe calls the apparently falsified corrections "shocking," particularly in light of Ashe's pride in the 2006 *Nature* paper. "I don't see how she would not hyperscrutinize anything that subsequently related to Aβ*56," he says.

After *Science* contacted Ashe, she separately posted to PubPeer a defense of some images Schrag had challenged in the *Nature* paper. She supplied portions of a few original, unpublished versions that do not show the apparent digital cut marks Schrag had detected in the published images. That suggests the markings were harmless digital artifacts. Yet the original images reveal something that Schrag and Selkoe find even more incriminating: unequivocal evidence that, despite the lack of obvious cut marks, multiple bands were copied and pasted from adjacent areas (see graphic, p. 361).

Schrag could find no innocent explanation for a 2-decade litany of oddities. In experiment after experiment using Western blots, microscopy, and other techniques, serious anomalies emerged. But he notes that he has not examined the original, uncropped, high-resolution images. Authors sometimes share those with researchers conducting similar work, although they usually ignore such requests, according to recent studies of data-sharing practices. Sharing agreements do not include access for independent misconduct detectives. Lesné and Ashe did not respond to a *Science* request for those images.

Questions about Lesné's work are not new. Cell biologist Denis Vivien, a senior scientist at Caen, co-authored five Lesné papers flagged by Schrag or Bik. Vivien defends the validity of those articles, but says he had reason to be wary of Lesné.

Toward the end of Lesné's time in France, Vivien says they worked together on a paper for *Nature Neuroscience* involving Aβ. During final revisions, he saw immunostaining images—in which antibodies detect proteins in tissue samples—that Lesné had provided. They looked dubious to Vivien, and he asked other students to replicate the findings. Their efforts failed. Vivien says he confronted Lesné, who denied wrongdoing. Although Vivien lacked "irrefutable proof" of misconduct, he withdrew the paper before publication "to preserve my scientific integrity," and broke off all contact with Lesné, he says. "We are never safe from a student who would like to deceive us and we must remain vigilant."

Schrag spot checked papers by Vivien or Ashe without Lesné. He found no anomalies—suggesting Vivien and Ashe were innocent of misconduct.

Yet senior scientists must balance the trust essential to fostering a protégé's independence with prudent verification, Wilcock says. If you sign off on images time after time, claim credit, speak publicly, and win awards for the work—as Ashe has done—you have to be sure it's right, she adds.

"Ashe obviously failed in that very serious duty" to ask tough questions and ensure the data's accuracy, Forsayeth says. "It was a major ethical lapse."

IN HIS WHISTLEBLOWER REPORT to NIH about Lesné's research, Schrag made its scope and stakes clear: "[This] dossier is a fraction of the anomalies easily visible on review of the publicly accessible data," he wrote. The suspect work "not only represents a substantial investment in [NIH] research support, but has been cited … thousands of times and thus has the potential to mislead an entire field of research."

The agency's reply, which Schrag shared with *Science*, noted that complaints deemed credible will go to the Department of Health and Human Services Office of Research Integrity (ORI) for review. That agency could then instruct grantee universities to investigate prior to a final ORI review, a process that can take years and remains confidential absent an official misconduct finding. To *Science*, NIH said it takes research misconduct seriously, but otherwise declined to comment.

In the fanfare around the Lesné-Ashe work, some Alzheimer's experts see a failure of skepticism, including by journals that published the work. After Schrag contacted *Nature*, *Science Signaling*, and five other journals about 13 papers co-authored by Lesné, a few are under investigation, according to emails he received from editors.

"There are very strong, legitimate questions," John Foley, editor of *Science Signaling*, later told *Science*. He says the journal has contacted authors and university officers of two papers from 2016 and 2017 for a response. It also recently issued expressions of concern about the articles.

A spokesperson for *Nature*, which publishes image integrity standards, says the journal takes concerns raised about its papers seriously, but otherwise had no comment. Days after an inquiry from *Science*, *Nature* published a note saying it was investigating Lesné's 2006 paper and advising caution about its results.

The Journal of Neuroscience stands out with five suspect Lesné papers. A journal spokesperson said it follows guidelines from the Committee on Publication Ethics to assess concerns, but otherwise had no comment.

"Journals and granting institutions don't know how to deal with image manipulation," Forsayeth says. "They're not subjecting images to sophisticated analysis, even though those tools are very widely available. It's not some magic skill. It's their job to do the gatekeeping."

Holden Thorp, editor-in-chief of the *Science* journals, said the journals have subjected images to increasing scrutiny, adding that "2017 would have been [near] the beginning of when more attention was being paid to this—not just for us, but across scientific publishing." He cited the Materials Design Analysis Reporting framework developed jointly by several publishers to improve data transparency and weed out image manipulation.

As federal agencies, universities, and journals quietly investigate Schrag's concerns, he decided to try to speed up the process by providing his findings to *Science*. He knows the move could have personal consequences. By calling out powerful agencies, journals, and scientists, Schrag might jeopardize grants and publications essential to his success.

But he says he felt an urgent need to go public about work that might mislead the field and slow the race to save lives. "You can cheat to get a paper. You can cheat to get a degree. You can cheat to get a grant. You can't cheat to cure a disease," he says. "Biology doesn't care."

Like other anti-Aβ efforts, toxic oligomer

ILLUSTRATION: N. DESAI/*SCIENCE*

Research backing experimental Alzheimer's drug was first target of suspicion

When Vanderbilt University physician and neuroscientist Matthew Schrag first grew suspicious of work underlying a major theory of Alzheimer's disease (see main story, p. 358), he was following a different trail. In August 2021, he provided analysis for a petition to the Food and Drug Administration (FDA), requesting that it pause two phase 3 clinical trials of Cassava Sciences's Alzheimer's drug Simufilam. The petition claimed some science behind the drug might be fraudulent, and the more than 1800 planned trial participants might see no benefits.

That month, Schrag submitted stinging reports to the National Institutes of Health (NIH) about 34 published papers by Cassava-linked scientists, describing "serious concerns of research misconduct." His findings, including possibly manipulated scientific images and suspect numerical data, challenge work supported by tens of millions of dollars in NIH funds. Some of the studies suggest Simufilam reinstates the shape and function of the protein filamin A, which Cassava claims causes Alzheimer's dementia when misfolded. (Other publications have reported on the FDA petition, but not Schrag's identity. *The Wall Street Journal* has reported that the U.S. Securities and Exchange Commission is also investigating Cassava.)

In February, FDA refused to pause the trials, calling the petition the wrong way to intervene, but said it might eventually take action. Independent image analysts and Alzheimer's experts who reviewed Schrag's findings at *Science*'s request generally agree with him.

Schrag's sleuthing implicates work by Cassava Senior Vice President Lindsay Burns, Hoau-Yan Wang of the City University of New York (CUNY), and Harvard University neurologist Steven Arnold. Wang and Arnold have advised Cassava, and Wang collaborated with the company for 15 years.

None agreed to answer questions from *Science*. Cassava CEO Remi Barbier also declined to answer questions or to name the company's current scientific advisers. He said in an email that Schrag's dossier is "generally consistent with prior allegations about our science … such allegations are false." Cassava hired investigators to review its work, provided "nearly 100,000 pages of documents to an alphabet soup of outside investigative agencies," and asked CUNY to investigate, he added. That effort "has yielded an important finding to date: there is no evidence of research misconduct." (CUNY says it takes allegations of misconduct seriously, but otherwise declined to comment because of its ongoing investigation.)

Last year, Schrag reached out to most of the journals that published questioned papers. Seven were retracted—including five by *PLOS ONE* in April. Three others received expressions of concern; in each case, the editors said they were awaiting completion of the CUNY investigation. In a few cases, the editors told him, reviews were underway.

Cassava has said editors of two suspect papers dismissed misconduct concerns. Last year, the editors of a 2005 *Neuroscience* paper co-authored by Wang, Burns, and others found no improper manipulation of Western blots, but said in an editorial note they would review any concerns from an "institutional investigation," apparently CUNY's probe. They did not respond to additional findings Schrag raised this year.

Another paper that purportedly validated science behind Simufilam—also by Wang, Burns, and colleagues—appeared in 2012 in *The Journal of Neuroscience*. In December 2021, the editors corrected one figure. Barbier said in a statement that they told him they had found no manipulation. But in January, after Schrag and others raised additional doubts, the editors issued an expression of concern—reserving judgment until CUNY completes its investigation.

Schrag received $18,000 from an attorney for short sellers behind the FDA petition, who profit if Cassava's value falls. Schrag, whose efforts were independent of Vanderbilt, says he worked hundreds of hours on the petition and independent research and he has never shorted Cassava stock or earned other money for efforts on that issue, or for similar work involving University of Minnesota, Twin Cities, neuroscientist Sylvain Lesné. (In either case, if federal authorities determine fraud occurred and demand a return of grant money, Schrag might be eligible to receive a portion of the funds.)

The most influential Cassava-related paper appeared in *The Journal of Clinical Investigation* in 2012. The authors—including Wang; Arnold; David Bennett, who leads a brain-tissue bank at Rush University; and his Rush colleague, neuroscientist Zoe Arvanitakis—linked insulin resistance to Alzheimer's and the formation of amyloid plaques. Cassava scientists say Simufilam lessens insulin resistance. They relied on a method in which dead brain tissue, frozen for a decade and then partially thawed and chopped, purportedly transmits nerve impulses.

Schrag and others say it contradicts basic neurobiology. Schrag adds that he could find no evidence that other investigators have replicated that result. (None of the authors agreed to be interviewed for this article.)

That paper supported the science behind Simufilam, Schrag says, "and spawned an entire field of research in Alzheimer's, 'diabetes of the brain.'" It has been cited more than 1500 times. Schrag sent the journal's editor his analysis of more than 15 suspect images. In an email that Schrag provided to *Science*, the editor said the journal had reviewed high-resolution versions of the images when they were originally submitted and declined to consider Schrag's findings. —*C.P.*

research has spawned no effective therapies. "Many companies have invested millions and millions of dollars, or even billions … to go after soluble Aβ [oligomers]. And that hasn't worked," says Daniel Alkon, president of the bioscience company Synaptogenix, who once directed neurologic research at NIH.

Schrag says oligomers might still play role in Alzheimer's. Following the *Nature* paper, other investigators connected combinations of oligomers to cognitive impairment in animals. "The wider story [of oligomers] potentially survives this one problem," Schrag says. "But it makes you pause and rethink the foundation of the story."

Selkoe adds that the broader amyloid hypothesis remains viable. "I hope that people will not become faint hearted as a result of what really looks like a very egregious example of malfeasance that's squarely in the Aβ oligomer field," he says. But if current phase 3 clinical trials of three drugs targeting amyloid oligomers all fail, he notes, "the Aβ hypothesis is very much under duress."

Selkoe's bigger worry, he says, is that the Lesné episode might further undercut public trust in science during a time of increasing skepticism and attacks. But scientists must show they can find and correct rare cases of apparent misconduct, he says. "We need to declare these examples and warn the world." ■

With reporting by Meagan Weiland. This story was supported by the *Science* Fund for Investigative Reporting.

INSIGHTS

PERSPECTIVES

GENOMICS

Reference genomes for conservation

High-quality reference genomes for non–model species can benefit conservation

By **Sadye Paez, Robert H. S. Kraus, Beth Shapiro, M. Thomas P. Gilbert, Erich D. Jarvis, the Vertebrate Genomes Project Conservation Group**

As of 2022, the International Union for Conservation of Nature (IUCN) Red List estimates that more than 32% of fungal, plant, and animal species are threatened with extinction. This sixth mass extinction is caused by the activities and expanding biomass of humans, necessitating a distinct name for this geological epoch—the Anthropocene (*1*). Human population growth and the vertebrate extinction rate (*2*) have been linearly correlated over the past 500 years (see the figure). For some species of conservation concern, documenting, informing, and mitigating this biodiversity loss has been helped by powerful genomic tools, including a reference assembly (*3*). Yet, currently, only a small fraction (<1%) of the ~35,500 species assessed as threatened with extinction have an available genome assembly, and to date, most are in draft form. It is proposed that conservation efforts can be enhanced by the production of high-quality reference genome assemblies.

Conservation genomics leverages genetic data, from individual loci to genomic-scale datasets, to aid preservation of species and population-level biodiversity. This includes using genomic data to measure effective population sizes, demographic history, and genetic diversity and to perform genetic manipulations pre- or postextinction. Many of these efforts have been conducted using first- and second-generation genome sequencing and assembly technologies with short reads, leading to sequence errors, structural errors, and missing sequences. Now, third-generation genome technologies—with improvements in longer read lengths, nucleotide accuracy, chromosomal maps, and assembly algorithms—have led to more complete assemblies (*4*). These high-quality assemblies have 10- to 200-fold improvements in quality metrics, including the amount of sequence assignable to chromosomes, genes fully assembled, and recovery of GC-rich regulatory regions (*4*). Method developments are also underway for generating complete and error-free genome assemblies (telomere to telomere) (*5*) of both maternal and paternal haplotypes. Given the extra computational and financial costs that such improved genomes incur, a legitimate question often asked is, what is the added value of these high-quality assemblies, beyond current draft genomes, for conservation?

Species must maintain a certain level of genetic diversity to adapt to various environmental changes and/or population decreases, whether natural or human driven. Genetic diversity or other genomic health assessments have historically drawn on polymerase chain reaction (PCR)–generated sequence data from DNA microsatellites, which are tandem repetitive sequences that tend to diverge at a higher rate compared with single-nucleotide variants. Typically, the greater the diversity in microsatellites among individuals, the healthier the population. Genomic health of a population can also be assessed by identifying changes in mutational load (the population frequency of deleterious alleles) and estimating lengths of runs of homozygosity (ROHs) (*6*). The accumulation of ROHs in small and inbred populations can fix or drive deleterious alleles to high frequencies. Identifying

Author affiliations and Vertebrate Genomes Project Conservation Group authors and affiliations are listed in the supplementary materials. Email: spaez@rockefeller.edu

PHOTO: STEPHEN BELCHER/MINDEN PICTURES

There were once ~30,000 kākāpō (*Strigops habroptilus*) on mainland New Zealand, but there are now only ~200 on nearby islands. High-quality reference genomes are aiding conservation breeding programs of these critically endangered parrots.

these and other signs of poor genomic health can provide a warning that a species is becoming critically endangered or identify specific populations that are fragile and deserving of focused conservation efforts. Although draft short read–based reference genomes have been successfully used for such analyses, third-generation reference genomes would lead to a comprehensive identification of microsatellites, mutational load, ROHs, and diverse segments of heterozygous variants (*6*) (see table S1).

Two examples of third-generation genome assemblies providing key information for conservation are those of the kākāpō (*Strigops habroptilus*) and vaquita (*Phocoena sinus*), which are both critically endangered (*4*) (see table S2). The kākāpō is a parrot endemic to New Zealand whose population once comprised ~30,000 individuals on the mainland. Human colonization circa 1360 CE and again in the 1800s reduced the population to 18 birds by 1977, but it is now recovering, with ~200 living on nearby islands. Analyses of a third-generation kākāpō genome and second-generation resequenced genomes from 49 individuals representing both extant and historical populations revealed that the surviving island population has had low genomic heterozygosity in long ROHs for the past 10,000 years, whereas the now-extinct mainland population did not (*6*). These findings affect conservation decision-making, whereby closely related individuals can now be bred with less concern for deleterious mutations, allowing a small population an opportunity to thrive.

The vaquita is a small porpoise endemic to the Gulf of California, Mexico, and is, at present, the world's most endangered marine mammal. Fewer than 19 individuals survive today, which is a reduction from a historical effective population size (N_e) of ~5000 (*7*) caused by bycatch in gillnets for shrimp and finfish over the past century. Inferred historical population analyses based on a third-generation vaquita genome assembly revealed that the species has had low genomic heterozygosity and a small N_e for the past ~250,000 years (*7*). This suggests that the vaquita, like the kākāpō, may have survived recent population declines because of effective long-term purging of deleterious mutations in the wild (*8*). Broadly, these findings indicate that low heterozygosity in a population will not always be detrimental if deleterious mutations have been purged and that introduction of individuals with higher heterozygosity but more deleterious alleles into a population with less deleterious alleles needs to be cautiously considered.

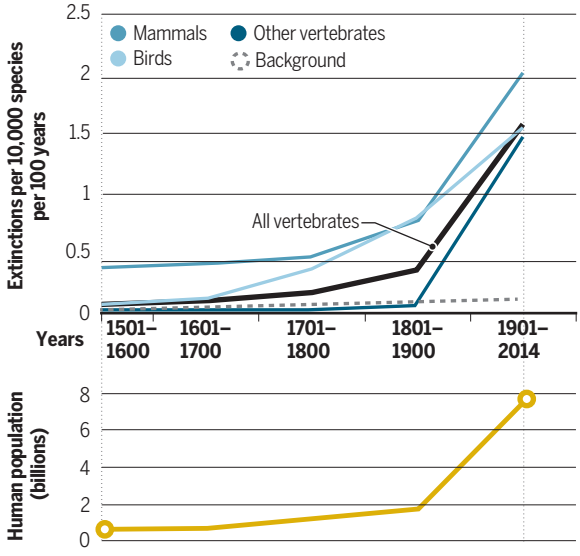
Genomic diversity can also be measured by counting and comparing structural variants (SVs), such as indels (insertions and deletions), inversions, chromosomal fusions, copy-number variations, and transposable elements. Identification of SVs is more straightforward in third-generation assemblies (*4*). SVs are increasingly appreciated as rich sources of adaptive polymorphism (*9*), representing conservation-relevant biological adaptations. For example, some SVs are adaptations to certain diseases, and thus selective breeding of individuals with these variants could potentially enhance population resistance to environmental changes.

Developing and implementing conservation management strategies often requires delineation of populations or species to distinguish between subspecies and cryptic species and, consequently, to differentiate conservation strategies. This was originally determined with DNA barcodes—that is, small fragments of DNA that are divergent between species, such as fragments of the organellular cytochrome oxidase subunit 1 and 16*S* and 12*S* ribosomal RNA genes—although this is increasingly complemented with analysis of whole-organelle sequences. The long-read approaches used in third-generation genome assemblies simplifies the reconstruction of entire organelle genomes (*4*). These organelle genomes have revealed repeat regions and gene duplications that were not assembled properly or were missed entirely in first- and second-generation genome assemblies (*4*). Additionally, these organelle genome assemblies, sometimes generated as a single-molecule sequence read, can be used for refined species delineation, phylogeography, and population studies; they also reduce problems that arise when, for example, mitochondrial genomes are confused with nuclear mitochondrial sequences (NUMTS).

Genetic rescue—which includes genetically informed translocations of a species from one geographical region to another, other breeding strategies, and more extreme interventions such as gene editing—aims to increase diversity or prevent the fixation of deleterious alleles by facilitating gene flow from one population to another. Although applications of gene editing have been mainly limited to agriculture, for example, to augment disease resistance in crops (*10*), proposed applications to conservation include improving a species' resistance to viral and bacterial infections or toxins and a species' capacity to adapt to anthropogenic and natural changes to their habitats, such as changes in temperature, salinity, or precipitation. Although these approaches are in the early stages of development and additional research is needed, third-generation genome assemblies may be critical to such efforts, for example, by

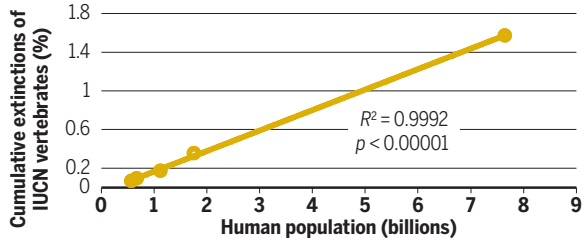
Vertebrate extinction rate and human population growth

The extinction rate of vertebrates, calculated according to historical and current records of the International Union for Conservation of Nature (IUCN) animal extinction list, is shown. All vertebrates as well as different species combinations are shown. The expected cumulative background extinction is based on geological extinction estimates between the fifth mass extinction (~65 million years ago) and 10,000 years ago. Graphs are modified from (*2*).



Human population correlates with extinction

There is a correlation between human population size and cumulative extinction of all vertebrates in the past 500 years; each point is a 100-year mean from the extinction data in the graph above.



GRAPHIC: N. DESAI/*SCIENCE*; (HUMAN POPULATION DATA) ECOLOGY.COM

better assembling translocations and having nearly all available sequences to determine potential off-target sites of genome editing.

Another potential benefit of third-generation genome assemblies to conservation will be for deextinction, such as resurrecting extinct traits in a living species or proxies of extinct species, for example, creating cold-adapted elephants using genomic diversity that evolved in the woolly mammoth (*11*). One approach is cloning using somatic cell nuclear transfer (SCNT), whereby nuclei of cells from extinct sublineages are transferred into enucleated oocytes that are then transplanted into a female. Preliminary reports indicate that this was successfully done for the Przewalski's horse and black-footed ferret with decades-old cryobanked cells, with the resulting clones still living in captivity (*12*). But this approach requires preserved living cells, which limits its application, and it is also not straightforward for egg-laying species such as birds and fishes. In these cases, gene editing of cells before SCNT or of early-stage embryos before egg formation might work better. But this approach requires knowledge of which edits to make. Contiguous and nearly complete genomes provide greater resolution to identify species-specific coding and regulatory sequences for gene editing.

In the absence of frozen cells, complete genome sequence data could also be used to create synthetic chromosomes and place them into viable cells, as was achieved by the Yeast 2.0 Project, which synthesized the entire genome of *Saccharomyces cerevisiae* (*13*). Although yeast genomes are 3 to 10% of the size of vertebrate genomes and the technology does not yet exist to synthesize larger genomes, this highlights the potential power for synthetic biology in deextinction efforts. In multicellular organisms, synthesized chromosomes could be placed in enucleated oocytes of another species, similar to the SCNT approach.

As genomic analyses and synthetic biology become components of conservation management, there are challenges to overcome, including developing approaches that consider other complex genome organizations, such as species with germline cells that have germline-specific chromosomes (e.g., lamprey and songbirds), and rearranged chromosomes during different developmental stages (e.g., single-cell ciliate protists) (*14*, *15*). Multicellular organisms also rely on microbial symbionts, some of which are inherited. High-quality genome assemblies for symbiotic microbes, such as those being developed by the Earth HoloGenome Initiative, are the crucial first step in incorporating this information into conservation management plans.

Addressing biodiversity loss is a complex problem that requires multifaceted solutions. Genomics can be an important component of conservation management. It is urgent that high-quality reference genome assemblies and cryopreserved cells be produced for endangered species now and, eventually, for all species. Waiting for technology improvements, policy changes, or outcomes of nongenomic efforts places too many species in peril. Generating high-quality genome assemblies from poorly preserved tissue or fossil remains (DNA can be extracted from samples up to 1 million years old in permafrost) is impossible because of the short lengths of surviving DNA molecules, highlighting the need for optimized cryopreservation of cells and tissues. When sex chromosomes exist, sequencing the heterogametic sex (e.g., males in mammals and females in birds) is preferable. Also, material should be preserved from multiple individuals so that information about population genetic diversity can be obtained. A notable and continuing challenge lies with the ethical, legal, and moral implications of translating genomic data to conservation. Coordination between scientists and other stakeholders is important, especially for access and benefit sharing of samples and the resulting digital sequence information with Indigenous Peoples and local communities. Genome assemblies by themselves, even if complete and error free, cannot fully address the ongoing sixth mass extinction. But high-quality reference genome assemblies are advantageous for pre- and postconservation management and monitoring with other strategies, such as preserving land, forest, and water reserves, and with other protections to the environment. ■

REFERENCES AND NOTES

1. C. N. Waters *et al.*, *Science* **351**, aad2622 (2016).
2. G. Ceballos *et al.*, *Sci. Adv.* **1**, e1400253 (2015).
3. P. Brandies, E. Peel, C. J. Hogg, K. Belov, *Genes* **10**, 846 (2019).
4. A. Rhie *et al.*, *Nature* **592**, 737 (2021).
5. S. Nurk *et al.*, *Science* **376**, 44 (2022).
6. N. Dussex *et al.*, *Cell Genom.* **1**, 100002 (2021).
7. P. A. Morin *et al.*, *Mol. Ecol. Resour.* **21**, 1008 (2021).
8. J. A. Robinson *et al.*, *Science* **376**, 635 (2022).
9. D. Weetman, L. S. Djogbenou, E. Lucas, *Curr. Opin. Insect Sci.* **27**, 82 (2018).
10. M. P. Phelps, L. W. Seeb, J. E. Seeb, *Conserv. Biol.* **34**, 54 (2020).
11. B. Shapiro, *Genome Biol.* **16**, 228 (2015).
12. Revive & Restore, Projects (2021); https://reviverestore.org/projects/.
13. I. S. Pretorius, J. D. Boeke, *FEMS Yeast Res.* **18**, foy032 (2018).
14. C. M. Kinsella *et al.*, *Nat. Commun.* **10**, 5468 (2019).
15. J. J. Smith *et al.*, *Nat. Genet.* **50**, 270 (2018).

ACKNOWLEDGMENTS

S.P. and R.H.S.K. contributed equally to this work.
See supplementary materials for full acknowledgments.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm8127

10.1126/science.abm8127

PLANETARY SCIENCE

In the glare of the Sun

Searches during twilight toward the Sun have found several asteroids near Venus' orbit

By **Scott S. Sheppard**

Asteroid surveys generally operate at night, mostly finding objects beyond Earth's orbit. This creates a blind spot because many near-Earth objects (NEOs) could be lurking in the sunlight interior to Earth's orbit. New telescopic surveys are braving the Sun's glare and searching for asteroids toward the Sun during twilight. These surveys have found many previously undiscovered asteroids interior to Earth, including the first asteroid with an orbit interior to Venus, 'Ayló'chaxnim (2020 AV2), and an asteroid with the shortest-known orbital period around the Sun, 2021 PH27 (*1*, *2*).

NEOs are classified into different dynamical types (see the figure). Starting from the most distant are the Amors, which approach Earth but do not cross Earth's orbit. Apollos cross Earth's orbit but have semimajor axes greater than that of Earth. Atens also cross Earth's orbit but have semimajor axes less than that of Earth. Atiras (also called Apohele) have orbits completely interior to Earth, and Vatiras have orbits completely interior to Venus, with 2020 AV2 being the first known.

NEOs have dynamically unstable orbits of ~10 million years. A reservoir must exist that replenishes the NEOs because their numbers have been in a steady state over the past few billion years (*3*). Most NEOs are likely dislodged objects from the main belt of asteroids between Mars and Jupiter (*4*–*6*). Physical observations show that NEOs are similar to main belt asteroids (MBAs), with a small fraction being dormant comets from the outer Solar System (*7*).

MBAs with orbital periods near whole number ratios with Jupiter's period are depleted, which indicates that these areas are dynamically unstable. Small MBAs continually move into these unstable regions

Earth and Planets Laboratory, Carnegie Institution for Science, Washington, DC 20015, USA.
Email: ssheppard@carnegiescience.edu

and are ejected from the main belt of asteroids through the Yarkovsky effect, which slowly changes an asteroid's orbit through the nonisotropic emitting of absorbed sunlight. The movement depends on the asteroid's rotation, size, albedo, and distance from the Sun. The smaller an asteroid is and the more sunlight it absorbs, the larger its movement.

Fewer Atiras should exist than the more-distant NEOs, and even fewer Vatiras, because it becomes harder and harder for an object to move inward past Earth's and then Venus' orbit. Random walks of a NEO's orbit through planetary gravitational interactions can make an Aten into an Atira and/or Vatira orbit and vice versa. Atiras should make up some 1.2% and Vatiras only 0.3% of the total NEO population coming from the main belt of asteroids (*4*). 2020 AV2 itself will spend only a few million years in a Vatira orbit before crossing Venus' orbit. Eventually, 2020 AV2 will either collide with or be tidally disrupted by one of the planets, disintegrate near the Sun, or be ejected from the inner Solar System.

Recent NEO models predict that there should be less than one Vatira of the roughly 1.5-km diameter of 2020 AV2 but many more smaller ones (*4*). Only a fraction of the sky has been searched where Vatira-like asteroids reside; however, because of the scattered light problem from the Sun, only the largest are observable. Finding a relatively large Vatira in the little area searched is somewhat unexpected, but small number statistics has caveats when trying to understand a whole population. Only a few asteroid surveys have imaged interior to Venus with published results (*8*, *9*), but the null results of others may be unpublished, making it hard to determine how much space interior to Venus has actually been well searched. This makes it difficult to get a true handle on Vatira discovery statistics.

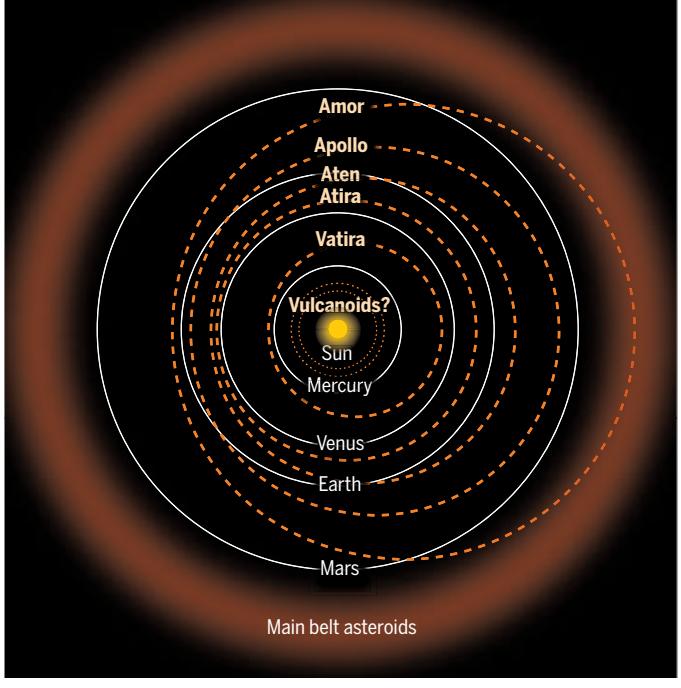
Recently, the asteroid with the smallest-known semimajor axis at 0.46 astronomical units (au) was found—2021 PH27 (*2*). Because of 2021 PH27's large eccentricity of 0.7, its orbit actually crosses both the orbits of Mercury and Venus, making it an Atira and not a Vatira asteroid. 2021 PH27 approaches so close to the Sun (0.13 au) that it has the strongest general relativity effects, at almost 1 arc min precession per century, of any known object in our Solar System, including Mercury. 2021 PH27's surface likely gets to 500°C, which is hot enough to melt lead. 2021 PH27 is also apparently ~1 km in size, which is relatively large. However, because the diameter of these interior asteroids is calculated with an assumed albedo and solar phase function, the actual diameters for both of these discoveries could be under 1 km (*10*). This would put them in a more-expected population and make them less of a statistical fluke.

Some Atiras and Vatiras could also have another source region besides the main belt of asteroids. These could be relatively stable inner reservoirs, such as objects in stable long-term resonances with Venus or Mercury, or even the hypothetical Vulcanoids (*8*). Vulcanoids are asteroids that could exist with orbits completely interior to Mercury's orbit that could be stable for billions of years, possibly forming there. Many exoplanets have been found closer to their host stars than the Vulcanoid region in our Solar System. Vulcanoids could also come from a random walk from the NEO population, but this would be very rare (*11*). Spacecraft observations of the near-Sun environment likely rule out Vulcanoids larger than ~5 km (*12*). Vulcanoids could be destabilized over long periods of time from Yarkovsky drift, collisions, and thermal fracturing so close to the Sun. Fewer than expected low-albedo, high-eccentricity NEOs with perihelia very close to the Sun have been found, likely because they fracture from the extreme thermal stresses (*13*). These dust-producing events may be the source of many meteor showers seen annually on Earth, like the Geminids meteor shower that occurs in mid-December from the shedding of material off the NEO Phaethon (*14*).

From NEO formation models and the current NEO survey efficiencies, more than 90% of planet-killer NEOs have been found (those larger than 1 km), although only about half of the city-killer NEOs are known (those larger than 140 meters). The last few unknown 1-km NEOs likely have orbits close to the Sun or high inclinations, which keep them away from the fields of the main NEO surveys. The 48-inch Zwicky Transient Facility telescope has found one Vatira and several Atira asteroids, making it one of the most prolific asteroid hunters interior to Earth. To combat twilight to find smaller asteroids, one can use a bigger telescope. Large telescopes usually do not have big fields of view to efficiently survey. The National Science Foundation's Blanco 4-meter telescope in Chile with the Dark Energy Camera (DECam) is an exception. A new search for asteroids hidden in plain twilight with DECam has found a few Atira asteroids, including 2021 PH27. These continuing twilight surveys are finally uncovering the population of small asteroids near the orbit of Venus. ■

Categorizing near-Earth objects

The different types of asteroids orbiting close to the Sun are classified on the basis of what planetary orbits they cross. The 'Ayló'chaxnim asteroid is the first Vatira type that has been observed. In principle, asteroids that only exist inside Mercury's orbit are also possible (the Vulcanoids) but have not yet been observed.



GRAPHIC: N. CARY/*SCIENCE*

REFERENCES AND NOTES

1. B. T. Bolin *et al.*, *MPEC* 2020-A99 (2020).
2. S. S. Sheppard *et al.*, *MPEC* 2021-Q41 (2021).
3. R. G. Strom *et al.*, *Res. Astron. Astrophys.* **15**, 407 (2015).
4. M. Granvik *et al.*, *Icarus* **312**, 181 (2018).
5. A. Morbidelli *et al.*, *Icarus* **340**, 113631 (2020).
6. A. W. Harris, P. W. Chodas, *Icarus* **365**, 114452 (2021).
7. R. P. Binzel *et al.*, *Icarus* **324**, 41 (2019).
8. P. Pokorný, M. J. Kuchner, S. S. Sheppard, *Planet. Sci. J.* **1**, 47 (2020).
9. Q. Ye *et al.*, *Astron. J.* **159**, 70 (2020).
10. A. Mainzer *et al.*, *Astrophys. J.* **752**, 110 (2012).
11. S. Greenstreet, H. Ngo, B. Gladman, *Icarus* **217**, 355 (2012).
12. A. J. Steffl, N. J. Cunningham, A. B. Shinn, D. D. Durda, S. A. Stern, *Icarus* **223**, 48 (2013).
13. Q. Ye, M. Granvik, *Astrophys. J.* **873**, 104 (2019).
14. D. Jewitt, J. Li, J. Agarwal, *Astrophys. J. Lett.* **771**, L36 (2013).

10.1126/science.abj9820

LASERS

To make a mirrorless laser

Periodic temporal modulation of a photonic crystal can be used to produce laser light

By **Daniele Faccio**[1] *and* **Ewan M. Wright**[2]

Inside any laser is a cavity with a "gain medium" that gives the laser its energy to emit light. A typical gain medium contains atoms that can be excited by using an external energy source and is sandwiched between a pair of mirrors. The mirrors impose a periodicity on the light inside the cavity—similar to how the length of a guitar string limits what musical notes can be played—and allows the medium to pack more energy into the light each time it passes through the gain medium. On page 425 of this issue, Lyubarov *et al.* (*1*) propose a radically new approach to making a laser in which the cavity is replaced by a medium with no mirrors. Instead, the optical properties of the medium are periodically modulated in time.

The laser device of Lyubarov *et al.* contains no mechanism for recirculating the light at all. Its operation relies on a slab of transparent material with a refractive index that varies periodically in time. Because the wavelength of light in a medium varies inversely with the refractive index—the shorter the wavelength, the higher the effective refractive index—the medium modulation produces an effect that is akin to periodically compressing light waves. Lyubarov *et al.* take advantage of this periodic temporal compression and show that it can be used to amplify light that will also be coherent, as is laser light.

Time-modulated systems and amplification of light from temporal modulation are not completely new. Although different research fields may trace the origins of these ideas back to different sources, they all connect to a series of ideas proposed in the mid-20th century. In 1970, physicist Gerald Moore explained how a temporally modulated yet otherwise empty cavity can lead to the creation of photons (*2*). This is commonly referred to as the dynamical Casimir effect, which was not experimentally verified in a superconducting circuit until 2011 (*3*). In general, any system that has a time-dependent parameter can exhibit some form of amplification, similar to how a child can increase the amplitude of a swing by shifting their weight periodically and strategically (*4*). A common feature of temporally periodic systems is a typical resonance frequency at which the energy transfers from the time-varying parameter—for example, the child periodically and strategically shifting their center of mass twice per period, first by bending their legs backward as they swing backward and then later extending their legs forward as they swing forward. In technical terms, the greatest amplification in energy for any periodic system occurs for light waves with a frequency equal to twice the parameter modulation frequency.

Researchers have been investigating how to use temporally modulated materials. For example, can such materials control the frequency of light or can magnet-free materials made of nonreciprocal elements be created in which light can only propagate in one direction (*5*)? One can draw on analogies between the spatial and temporal cases of photon modulation in crystals to understand the physics of periodic time crystals. Spatial photonic crystals are crystal-like materials with periodic structures that modulate the propagation of light (*6*). These crystals behave for light in a way similar to what atomic crystals do for electrons, in that they lead to the formation of periodic bandgap structures—forbidden "gaps" in the frequency range where the propagation of waves is strongly suppressed. This suppression of waves occurs when the wave vector of the light is equal to half of the periodicity of the modulation and can be used to confine light, similar to the mirrors of a standard laser.

> **"...the 'transparent' atom acts as a conduit for energy transfer..."**

To observe the formation of "gaps" with a temporal modulation, one may use a block of material that can change its refractive index with the right periodicity. For such a system, one can expect a bandgap where the frequency is equal to half of the temporal modulation frequency of the material. When this happens, the energy of the system is no longer conserved, which allows its energy to be amplified. Although this was known for a wave propagating inside a periodic time crystal, Lyubarov *et al.* provide detailed classical and quantum models for an atom placed inside such a crystal. When stimulated with a flash of light, an atom inside the medium remains essentially in a so-called "transparent" state in which there are an equal number of electrons in the ground and excited states. In this state of balance, the stimulation causes the atom to absorb and emit equal amounts of light. The periodic modulation can then lead to exponential amplification for the emitted light with a narrowed spectrum that is characteristic of a laser beam. In this view, the "transparent" atom acts as a conduit for energy transfer between the periodic modulation of the medium and the emitted light. Moreover, this behavior does not appear to depend on the specific initial excitation of the atom. The initial stimulation with a flash of light can be at a substantially different frequency from the bandgap frequency as long as the light is emitted across a broad range of frequencies. Eventually, the exponential amplification at the bandgap will take over and pin the system to emission at the resonant frequency at half the modulation frequency.

Although the periodic time crystal laser does not rely on cavity mirrors or a gain medium, it does rely on a modulation of the medium that needs to be extremely fast because of the resonance condition, with the exponential amplification rate depending on the amplitude of the modulation. Typical photonic materials exhibit small modulations of the refractive index at the femtosecond or picosecond time scales required for lasers at visible to terahertz wavelengths. Recent progress in so-called epsilon-near-zero or index-near-zero materials offers a possibility for ultrafast switching of the medium with near-unity refractive index modulation (*7*, *8*), but this typically also has large losses that may make it harder to achieve laser-like behavior.

The mechanism presented by Lyubarov *et al.* may also be applied for producing light at much longer wavelengths by transducing different forms of energy into electromagnetic radiation. For example, a periodic temporal mechanical modulation in the form of a periodic pressure applied to the medium could be used to amplify electromagnetic waves, akin to the original proposal by Moore, albeit not with a cavity but through a photonic time crystal—more than half a century after the idea was first proposed. ■

[1]School of Physics and Astronomy, University of Glasgow, Glasgow, UK. [2]Wyant College of Optical Sciences, University of Arizona, Tucson, AZ 85721, USA.
Email: daniele.faccio@glasgow.ac.uk; ewan@optics.arizona.edu

REFERENCES AND NOTES

1. M. Lyubarov *et al.*, *Science* **377**, 425 (2022).
2. G. T. Moore, *J. Math. Phys.* **11**, 2679 (1970).
3. C. M. Wilson *et al.*, *Nature* **479**, 376 (2011).
4. P. D. Nation *et al.*, *Rev. Mod. Phys.* **84**, 1 (2012).
5. E. Galiffi *et al.*, *Adv. Photonics* **4**, 014002 (2022).
6. J. D. Joannopoulos *et al.*, *Molding the Flow of Light* (Princeton Univ. Press, ed. 2, 2008).
7. M. Z. Alam, I. De Leon, R. W. Boyd, *Science* **352**, 795 (2016).
8. L. Caspani *et al.*, *Phys. Rev. Lett.* **116**, 233901 (2016).

10.1126/science.abq5012

PLASTICS

Improving catalysis by moving water

The conversion of gases into building blocks for synthesizing plastics is enhanced

By **Mingyue Ding** *and* **Yanfei Xu**

Light olefins, which include commercially important chemicals such as ethylene, propylene, and butylene, are used to fabricate a wide range of plastics and synthetic fibers (*1*). Because they are typically produced from petroleum, their production is intrinsically linked to problems stemming from petroleum extraction and processing. Researchers have been exploring alternative routes for the synthesis of light olefins and have found success in making light olefins from syngas, a mixture of carbon monoxide (CO) and hydrogen (H_2) that can be derived from not only coal and natural gas but also non–fossil fuel–based biomass (*2*–*4*). Although considerable progress has been made in syngas conversion, the ability to produce light olefins from syngas remains limited. On page 406 of this issue, Fang *et al.* (*5*) present a simple and effective method to boost the conversion of syngas to light olefins.

During syngas conversion, the carbon-carbon coupling that is catalyzed by conventional metals or carbide-based metals follows a polymerization step in which the carbon chain grows without control. This produces hydrocarbon products with a wide range of carbon numbers, which means a poor selectivity for light olefins (*6*). Designing a catalyst with excellent selectivity for light olefins is challenging. An iron-based catalyst, in which iron is supported on aluminum oxide ($Fe/\alpha\text{-}Al_2O_3$) exhibited 53% selectivity for light olefins with 80% CO conversion (*2*). Cobalt manganese (CoMn) as a catalyst showed a slightly better 61% selectivity for light olefins with 32% CO conversion (*3*). In the other catalyst system, called the oxide-zeolite route, the carbon-carbon coupling proceeds in the confined zeolite micropore, and thus it is easy to control the carbon number of hydrocarbon products by adjusting zeolite pore size. A superior selectivity of 80% for light olefins with 17% CO conversion was achieved over a mixture of partially reduced oxide and mesoporous silicoaluminophosphate zeolite ($ZnCrO_x$/MSAPO) (*4*).

Water is a by-product of the CO hydrogenation reaction and can limit the efficiency of syngas to light olefins by covering active sites or inducing undesirable reactions. Fang *et al.* report a catalyst that is a mixture of a CoMn catalyst with polydivinylbenzene (PDVB). The CoMn/PDVB mixture possesses the advantages of the CoMn catalyst—its high selectivity for light olefins and mild reaction conditions—whereas the PDVB acts as hydrophobic water-conduction channels to move water away from the CoMn catalyst (see the figure). This combination leads to a substantial increase in CO conversion, at 64%, and a good selectivity for light olefins in hydrocarbon products, at 71%, achieved under mild reaction conditions of 250°C and 0.1 MPa.

After removing the PDVB in the CoMn/PDVB mixture after use, the remaining CoMn exhibited catalytic activity similar to that of fresh CoMn. This observation suggests that hydrophobic PDVB does not change the structure of CoMn during the reaction but instead influences the water-sorption equilibrium on the CoMn surface through its hydrophobicity. To support this interpretation, Fang *et al.* examined CoMn during a reaction in which water is added to the syngas feed. After the deliberate injection of water, the CO conversion over CoMn decreased from 34 to ~8%, whereas CO conversion over CoMn/PDVB only decreased slightly, from 64 to ~55%. Further pulse experiments and in situ diffuse reflectance infrared Fourier transform spectroscopy revealed the hindrance of CO adsorption by competition with water on the CoMn surface. The addition of hydrophobic PDVB to CoMn efficiently shifts the sorption equilibrium of water, thereby reducing the negative effect of water on the adsorption and conversion of CO molecules.

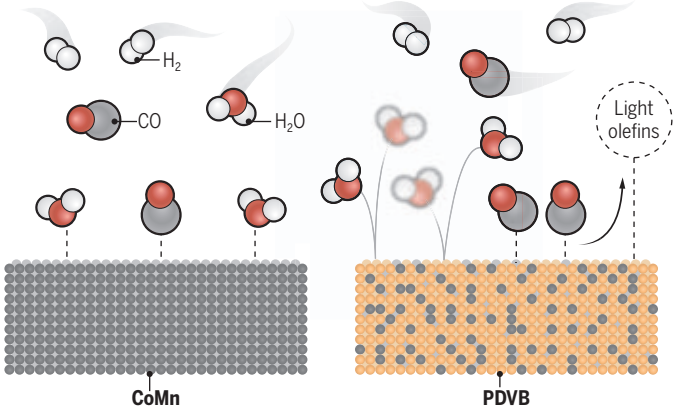
Fang *et al.* also performed theoretical simulations to study the effect of channel wettability on water diffusion. The hydrophilic channel interacts with water molecules and slows down their diffusion, whereas the weak interaction between the hydrophobic channel and water molecules accelerates the escape of water. The simulation results suggest that even though the water-adsorbed region of the CoMn catalyst and the channels are separated from each other, more water molecules escape from the hydrophobic channel than the hydrophilic one.

Mixing a hydrophobic promoter with a CoMn catalyst is a simple but effective strategy to enhance the conversion of syngas to light olefins. Fang *et al.* provide a method to enhance the conversion efficiency without influencing the selectivity for target products. The improved efficiency will reduce the manufacturing cost that is incurred by the undesirable but necessary repeated reaction of unreacted syngas during production. The use of a hydrophobic promoter to accelerate water escape and thereby expose more active sites on the catalyst for reactants may be applicable in other water-restricted reactions, such as carbon dioxide (CO_2) catalytic hydrogenation, which can be used to produce fossil fuel–free fuel.

Before the findings of Fang *et al.*, there had been other hydrophobization strategies,

Reducing the negative effect of water on light olefin catalysis

The competitive adsorption of water (H_2O) and carbon monoxide (CO) on a cobalt manganese (CoMn) catalyst limits syngas conversion (left). Hydrophobic polydivinylbenzene (PDVB) acts as water-conduction channels and accelerates the diffusion of water, thereby exposing more active sites on the CoMn catalyst for converting syngas to light olefins (right).



GRAPHIC: N. DESAI/*SCIENCE*

School of Power and Mechanical Engineering, The Institute of Technological Sciences, Wuhan University, Wuhan 430072, China. Email: dingmy@whu.edu.cn

such as chemically modifying the catalyst surface (*7*), which reduces CO_2 selectivity by suppressing the water-gas shift reaction (i.e., $CO + H_2O \rightleftharpoons CO_2 + H_2$). Both the mixing and the chemically modifying strategies enhance the conversion of CO to hydrocarbon products, but the two approaches differ because of the difference in the distance between the hydrophobic entity and the catalytic active sites in each case.

For the chemically modifying route, the hydrophobic compound coats the catalytic active sites, inhibiting the readsorption of water on the catalyst. In this case, most of the water produced during the reaction can escape without participating in the water-gas shift reaction. The water-gas shift reaction deviates away from equilibrium, giving a low CO_2 selectivity of 13% with a CO conversion rate of 56%. For the mixing route, the CO_2 selectivity over CoMn/PDVB is near 50%, implying that the water-gas shift reaction is near equilibrium. The hydrophobic compound and the catalytic active sites are separated from each other, and the distance between them is nanometers to microns. In this case, most of the water produced on CoMn participates in the water-gas shift reaction before diffusing to the hydrophobic promoter. Fang *et al.* report that the CoMn catalyst is water sensitive, and the small amount of water that is not involved in the water-gas shift reaction can hinder the adsorption of CO on CoMn and inhibit CO conversion. The addition of hydrophobic PDVB accelerates the diffusion of this unused water, exposing more active sites on the CoMn catalyst for the adsorption and conversion of CO.

Decreasing the CO_2 release during syngas conversion can reduce costs and enhance the productivity of the target product (*8*, *9*). As countries propose carbon-neutral goals to address climate change, carbon emission reduction in the chemical industry is imperative. Combining the hydrophobic strategies of mixing or chemical modification to develop new catalysts with low CO_2 selectivity and high CO conversion will unlock the full potential of this process to produce valuable chemicals from syngas. ■

REFERENCES AND NOTES

1. X. Pan, F. Jiao, D. Miao, X. Bao, *Chem. Rev.* **121**, 6588 (2021).
2. H. M. Torres Galvis *et al.*, *Science* **335**, 835 (2012).
3. L. Zhong *et al.*, *Nature* **538**, 84 (2016).
4. F. Jiao *et al.*, *Science* **351**, 1065 (2016).
5. W. Fang *et al.*, *Science* **377**, 406 (2022).
6. W. Zhou *et al.*, *Chem. Soc. Rev.* **48**, 3193 (2019).
7. Y. Xu *et al.*, *Science* **371**, 610 (2021).
8. P. Wang *et al.*, *Sci. Adv.* **4**, eaau2947 (2018).
9. Y. Xu, X. Li, M. Ding, *Chem* **7**, 1977 (2021).

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (21978225 and U21A20317).

10.1126/science.adc9414

PLANT BIOLOGY

The quest for more food

Rice yield is increased by boosting nitrogen uptake and photosynthesis

By **Steven Kelly**

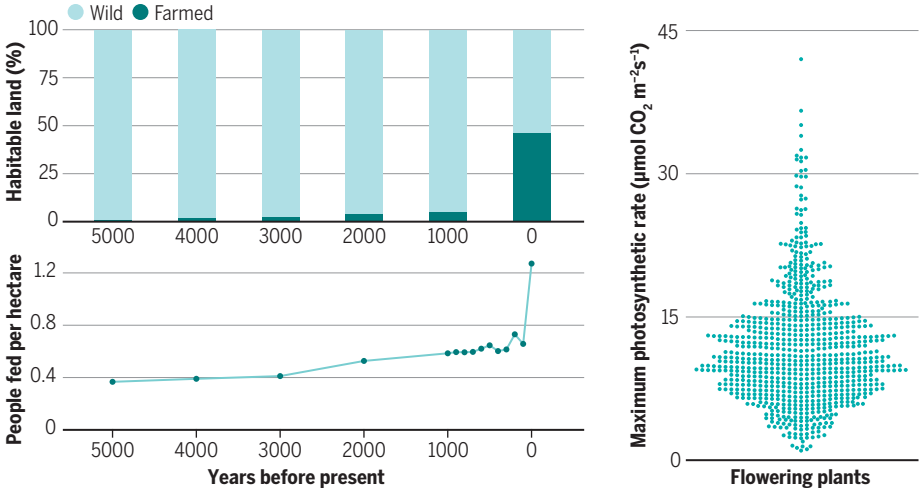
Enhancing photosynthesis is regarded as one of the most promising avenues for increasing crop yield (*1*). Accordingly, substantial focus has been drawn to this challenge, with several breakthroughs holding great potential (*2*). A common feature of these successes is that they have targeted metabolic processes, altering the rate and/or the path of metabolite flow through the plant to achieve higher rates of photosynthesis. However, on page 386 of this issue, Wei *et al.* (*3*) report an alternative approach. They show that photosynthesis and yield can be improved in rice by overexpressing a transcriptional regulator that promotes the expression of yield-associated genes. This highlights that there is a substantial latent capacity for enhancing photosynthesis hidden in the genomes of plants. Moreover, this latent capacity is present in abundance even in plants subjected to thousands of years of improvement through plant breeding.

Planet Earth is home to more than 6000 species of mammal (*4*), 300,000 species of plant (*5*), and 5,000,000 species of insect (*6*). Evolution by natural selection took half a billion years to craft this natural diversity and distribute it over the ~104 million km^2 of habitable land. However, in just the past 5000 years, humanity has replaced ~50% of this wilderness with agriculture (*7*) (see the figure). This rapid destruction of the natural world is causing the sixth global mass extinction (*8*), with current rates of extinction unseen since an asteroid wiped out the dinosaurs (*8*). Moreover, the conversion of complex ecosystems into agricultural systems has released billions of tons of CO_2 into the atmosphere, accelerating climate change, reducing the capacity to store CO_2, and placing further pressure on the natural world (*9*)—all in the quest for food.

Although the main way in which humanity has increased food production has been through the expansion of agricultural land, this expansion has been mitigated by scientific technological development. For example, plant breeders have improved the shape and form of domesticated plants so that their canopy intercepts almost all of the light before it hits the ground and so that the largest possible fraction of the carbon captured over the lifetime of the

Changing landscapes

The proportion of total habitable land that is wild versus farmed has decreased over time (*7*) (top left). Additionally, the amount of people fed per hectare has increased, owing to technological improvements (bottom left). However, this is not enough to feed the human population, so further yield gains are needed. Photosynthetic rate is highly variable in different plant species (*11*) (right), indicating that improvements to photosynthesis, and thus yield, are possible.



GRAPHIC: K. FRANKLIN/*SCIENCE*

plant ends up in the grain (*1*). These masterworks of plant engineering, coupled with a near-limitless supply of fertilizer from the Haber-Bosch process (*10*), have enabled humanity to feed more people per unit of land area than ever before (*7*). Today, it takes only ~30% as much land to feed one person as it did 5000 years ago. The problem is that there are 138 times as many people alive today as there were 5000 years ago. Thus, this improvement in land use is not enough.

Because photosynthesis provides all of the carbon and energy that plants use to grow, enhancing photosynthesis holds great potential to increase crop growth and yield. Notably, there is extensive natural diversity in the rates of photosynthesis between plants (*11*). Some plants have evolved high rates of photosynthesis to optimize growth and reproduction in the shortest time possible. However, some plants barely grow at all, only reproducing once a century or less. Our ancestors did not choose plants to domesticate based on how good they were at photosynthesis. They domesticated plants because they tasted good, did not poison them (mostly), and provided a reliable source of easy-access nutrition. By chance, some of these plants, such as *Zea mays* (maize), are extremely efficient at photosynthesis. However others, such as rice, are just average. This disparity has inspired researchers to ask how photosynthesis can be improved in average plants like rice. Success has come from thinking carefully about how photosynthesis works and what its limitations and bottlenecks are. Examples of these successes include speeding up plant responses to fluctuating light (*12*), the creation of metabolic bypasses to minimize energy losses (*13*), and enhancing the capacity of the engine of photosynthesis—the chloroplast (*14*).

"Our ancestors did not choose plants to domesticate based on how good they were at photosynthesis."

Wei *et al.* found a new way to improve photosynthesis. They discovered it by asking, what does rice do when challenged with stressful growth conditions? They noticed that expression of the OsDREB1C transcription factor was up-regulated when rice plants were grown under low-nitrogen conditions. When they overexpressed OsDREB1C in rice, they increased nitrogen uptake and boosted photosynthesis, the net result being a 12 to 40% increase in yield in rice plants grown in the field. They also found that overexpressing OsDREB1C in wheat and the model plant *Arabidopsis thaliana* similarly improved yield (or biomass) by more than 10%.

Pinning down the mechanism that gave rise to this yield increase was challenging because there were thousands of genes whose expression changed in response to the overexpression of OsDREB1C. However, chromatin immunoprecipitation coupled to RNA sequencing revealed that key nitrogen assimilation enzymes and transporters were direct downstream targets of OsDREB1C. This enabled plants to take up more nitrogen from their environment and support a higher rate of photosynthesis, faster growth, and more yield. Ultimately, this work revealed that there is a gold mine of potential for enhancing growth and yield that can be uncovered by learning how plants respond to their environment.

If there is to be both a sustainable future for humanity and more space for wildlife, then complex issues, such as reducing food waste and altering the relative use of animal and plant proteins, need to be addressed. However, even if these challenging goals can be achieved, averting the sixth global mass extinction will require producing a larger quantity of food than has ever been produced in human history, on much less land than is being used today. For this reason, increasing photosynthesis, and consequently the amount of food that can be produced per unit of land area, is one of the most important objectives for the sustainable future of the planet. ■

Department of Plant Sciences, University of Oxford, Oxford, UK. Email: steven.kelly@plants.ox.ac.uk

REFERENCES AND NOTES

1. S. P. Long, A. Marshall-Colon, X.-G. Zhu, *Cell* **161**, 56 (2015).
2. A. J. Simkin, P. E. López-Calcagno, C. A. Raines, *J. Exp. Bot.* **70**, 1119 (2019).
3. S. Wei *et al.*, *Science* **377**, 386 (2022).
4. C. J. Burgin, J. P. Colella, P. L. Kahn, N. S. Upham, *J. Mammal.* **99**, 1 (2018).
5. M. J. M. Christenhusz, J. W. Byng, *Phytotaxa* **261**, 201 (2016).
6. N. E. Stork, *Annu. Rev. Entomol.* **63**, 31 (2018).
7. K. Klein Goldewijk, A. Beusen, J. Doelman, E. Stehfest, *Earth Syst. Sci. Data* **9**, 927 (2017).
8. R. H. Cowie, P. Bouchet, B. Fontaine, *Biol. Rev.* **97**, 640 (2022).
9. S. R. Weiskopf *et al.*, *Sci. Total Environ.* **733**, 137782 (2020).
10. V. Smil, *Ambio* **31**, 126 (2002).
11. J. Gago *et al.*, *Trends Plant Sci.* **24**, 947 (2019).
12. J. Kromdijk *et al.*, *Science* **354**, 857 (2016).
13. R. Kebeish *et al.*, *Nat. Biotechnol.* **25**, 593 (2007).
14. X. Li *et al.*, *Commun. Biol.* **3**, 151 (2020).

ACKNOWLEDGMENTS

S.K. is a cofounder of Wild Bioscience LTD and provides consultancy to the company.

10.1126/science.add3882

MEDICINE

One step closer to cancer nanomedicine

High-throughput tool uncovers links between cell signaling and nanomaterial uptake

By **Jessica O. Winter**[1,2]

The promise of chemotherapeutic nanomedicine has tantalized clinicians and patients for decades. Nanoparticles (NPs) can directly target tumor cells, which would reduce the amount of chemotherapy administered and its systemic toxicity, increasing patient quality of life and extending utility of therapies with lifetime dosing limits. However, these hopes remain largely unrealized. Liposomal drug carriers, which make up nearly all clinically approved nanomedicines, have not extended overall patient survival compared with treatment with the drugs alone (*1*). These failures have been attributed to poor delivery to target cells (*2*) because NPs must first traverse a series of biological barriers (*3*). Although nanocarrier composition, surface chemistry, size, and shape have been optimized to promote cell entry, progress has been confounded by heterogeneity in cell uptake signaling (*4*). On page 384 of this issue, Boehnke *et al.* (*5*) uncover the reciprocal relationship between NP material properties and cell internalization using nanoPRISM, a high-throughput screening approach.

The nanoPRISM technology uses the profiling relative inhibition simultaneously in mixtures (PRISM) (*6*) method to generate a screening library of ~500 cancer cell lines that are barcoded with distinct DNA sequences that permit identification of cells with high-throughput genomic sequencing. This cell library is combined with a panel of 35 different fluorescently labeled NPs with varying core compositions, surface chemistries, and diameters to identify synergistic interactions for cell uptake. PRISM-tagged cells are separated

[1]William G. Lowrie Department of Chemical and Biomolecular Engineering, Ohio State University, Columbus, OH, USA. [2]Department of Biomedical Engineering, Ohio State University, Columbus, OH, USA. Email: winter.63@osu.edu

into four groups according to uptake level, and their DNA is sequenced to identify them and screen for key drivers of NP internalization that can be attributed to either NP characteristics or cell signaling.

Boehnke *et al*. compared the uptake efficiency of NPs conjugated to antibodies targeting epidermal growth factor receptor (EGFR) versus EGFR antibodies alone in cell lines that overexpress this receptor. NanoPRISM revealed differences in cellular uptake, most likely resulting from the steric hindrance of NP conjugation. These results suggest that nanoPRISM may be suitable for evaluating antibody-drug conjugates (ADCs), a growing therapeutic category.

Boehnke *et al*. also use nanoPRISM to interrogate NPs with compositions most commonly applied to nanomedicine: spherical liposomes made of lipid bilayers and solid lipid and polymer NPs consisting of disordered, spherical lipid or polymer aggregates. They also examine NPs with or without polyethylene glycol (PEG) modification, which is used to reduce systemic uptake and improve circulation time (*7*). They find that NP core composition is a primary determinant in cellular uptake. This unexpected finding upends years of work on modulating NP surface chemistries to alter protein adsorption patterns and subsequent cell adhesion (*8*). Although cells first detect NPs through their surface chemistry, the findings of Boehnke *et al*. support early studies that showed that NP stiffness and deformability, which are dictated by core composition, are stronger modulators of the uptake process (*9*).

The power of the nanoPRISM method is further illustrated by combining these findings with the Cancer Cell Line Encyclopedia, which quantifies mutational genomic signatures of common cancer cell lines. Boehnke *et al*. identify genomic signatures and signaling networks most correlated with NP internalization. Many of the results are not surprising, such as involvement of the solute carrier (SLC) transporter or adenosine triphosphate (ATP)–binding cassette (ABC) families, which have previously been implicated in NP cellular entry and transport. The nanoPRISM screens also highlight gene networks associated with the plasma membrane and extracellular matrix that contribute to NP cellular entry processes (see the figure).

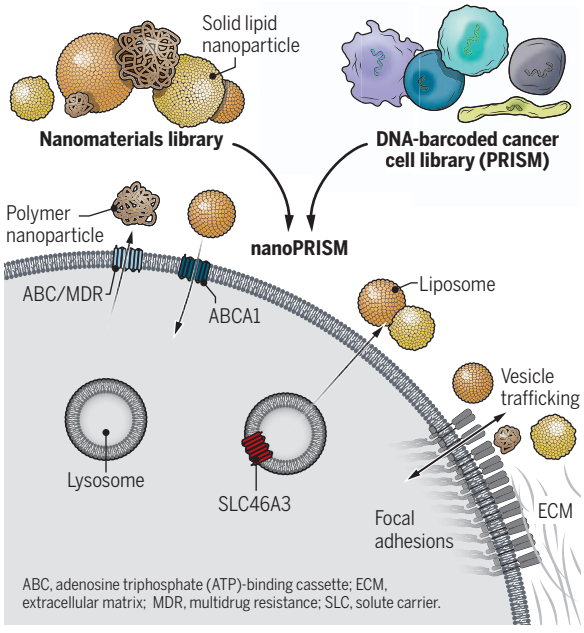
However, the nanoPRISM method also reveals involvement of an understudied gene that has not been associated with NP internalization: *SLC46A3*. This encodes a lysosomal transmembrane protein linked to lipid catabolism (*10*) that influences lysosomal trafficking of ADCs (*11*). Expression of SLC46A3 negatively regulated liposomal and solid lipid NP cellular uptake, whereas polymer NPs that lack lipids were unaffected. SLC46A3 association with lipid-based NPs was evidenced even when NP surfaces were coated with nonlipid molecules. This further indicates the importance of NP core composition in cellular uptake processes and also suggests that cells can detect core composition through surface coatings, which better resemble a porous net than a wall. This could have important implications for predicting the efficacy of nucleic acid vaccines and therapies that use lipid-based carriers, such as COVID-19 mRNA vaccines. For example, SLC46A3 biomarker testing could be implemented to identify patients most likely to respond to lipid-based nanotherapeutics.

The results of the nanoPRISM screens are also confirmed in animal models, indicating that this technique could be used to identify the most promising formulations for downstream analysis, reducing preclinical animal testing demands. Such high-throughput approaches are critical to the rapid advancement of cancer nanomedicine, because US and European regulatory agencies have not established criteria for nanomedicine approval based on similarity to an existing product (*12*). Given the long timeline for drug development, which can span a decade or more, technologies to safely accelerate this process are desirable.

The nanoPRISM method represents a substantial advance over the less rigorous and qualitative studies of NP internalization that characterized the early years of the field. Studies that examined a few NP properties in a single cell line could not capture the complexities of NP cell entry. Combined with machine learning and iterative simulation and materials synthesis approaches, nanoPRISM could enable screening for nanomaterials that target specific cell types, similar to current biopanning methods for peptides or the systematic evolution of ligands by exponential enrichment (SELEX) method of aptamer discovery (*13*). Although the study of Boehnke *et al*. examines only 35 different NPs, additional nanomaterials could be added to the library, such as inorganic NPs (such as gold, silica, and carbon) and materials with complex geometries (such as DNA origamis). A limitation of nanoPRISM is its focus on cellular entry, the last step of the biodistribution process. However, it is easy to envision expanding this approach beyond cell uptake to study the relationship between NP material properties and gene expression in cell adhesion and trafficking. Additionally, with the template provided by Boehnke *et al*., similar methods could be integrated with microfluidics, organ-on-a-chip, or tumor organoid cultures to model other delivery barriers, such as circulation, extravasation, and tissue diffusion. Thus, the nanoPRISM approach could catalyze rapid materials optimization, accelerating nanocarrier design and bringing the promise of cancer nanomedicine closer to reality. ■

Signatures of cellular uptake

The nanoPRISM method combines cell and nanomaterial libraries to identify signatures associated with cellular internalization. The ABC and SLC protein families regulate uptake of lipid-based and polymer nanoparticles differentially, whereas vesicular trafficking, ECM, and focal adhesion pathways affected all types of nanoparticles. Core composition, not surface chemistry, was the strongest regulator of uptake behavior.



ABC, adenosine triphosphate (ATP)-binding cassette; ECM, extracellular matrix; MDR, multidrug resistance; SLC, solute carrier.

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

REFERENCES AND NOTES

1. G. H. Petersen, S. K. Alzghari, W. Chee, S. S. Sankari, N. M. La-Beck, *J. Control. Release* **232**, 255 (2016).
2. S. Wilhelm *et al*., *Nat. Rev. Mater.* **1**, 16014 (2016).
3. S. Barua, S. Mitragotri, *Nano Today* **9**, 223 (2014).
4. B. D. Chithrani, A. A. Ghazani, W. C. W. Chan, *Nano Lett.* **6**, 662 (2006).
5. N. Boehnke *et al*., *Science* **377**, eabm5551 (2022).
6. C. Yu *et al*., *Nat. Biotechnol.* **34**, 419 (2016).
7. M. Eugene, *Cell. Mol. Biol.* **50**, 209 (2004).
8. A. Albanese *et al*., *ACS Nano* **8**, 5515 (2014).
9. X. Sun *et al*., *Biomacromolecules* **6**, 2541 (2005).
10. J.-H. Kim *et al*., *Nat. Commun.* **12**, 290 (2021).
11. K. J. Hamblett *et al*., *Cancer Res.* **75**, 5329 (2015).
12. S. Soares, J. Sousa, A. Pais, C. Vitorino, *Front Chem.* **6**, 360 (2018).
13. C. Tuerk, L. Gold, *Science* **249**, 505 (1990).

10.1126/science.add3666

POLICY FORUM

WATER GOVERNANCE

What will it take to stabilize the Colorado River?

A continuation of the current 23-year-long drought will require difficult decisions to prevent further decline

By **Kevin G. Wheeler[1,2], Brad Udall[3], Jian Wang[4], Eric Kuhn[5], Homa Salehabadi[4], John C. Schmidt[6]**

The Colorado River supplies water to more than 40 million inhabitants in the southwestern United States and northwestern Mexico. A basin-wide water supply crisis is occurring because of decreased watershed runoff caused by a warming climate and legal and water management policies that allow systematic overuse. By the end of 2022, combined storage in Lake Powell and Lake Mead, the two largest reservoirs in the United States, will have declined from 95% full in 2000 to approximately 25% full. If this "Millennium Drought" persists, then stabilizing reservoir levels to avoid severe outcomes will require reducing water use to match diminished runoff. With a process underway to renegotiate interstate and international agreements on consumptive uses of the river, we describe a promising new management approach based on combined storage of both reservoirs, rather than just Lake Mead as currently used, to trigger consumptive use reductions to the Lower Basin and Mexico.

Since 2000, the average annual natural flows (that which would exist without human interventions) into lakes Powell and Mead have been almost 20% below the 20th-century average (*1*, *2*). As a result of these unprecedented low flows and insufficient management adaptations (*3*, *4*), 5-year projections by the US Bureau of Reclamation(Reclamation) suggest that Lake Powell, created by Glen Canyon Dam, has a one in four chance of falling below the minimum elevation necessary to produce hydropower. Storage downstream in Lake Mead, created by Hoover Dam, has a two in five chance of falling to its most severe management condition, which forces large reductions on downstream users (*5*).

Municipalities of Los Angeles, San Diego, Phoenix, Tucson, Las Vegas, Denver, Salt Lake City, Albuquerque, and Tijuana rely heavily on the river for their water supplies. About 70% of the water is used to irrigate nearly 5.7 million acres (2.3 million hectares) of agriculture. The basin is home to 30 recognized Native American Tribes that hold senior legal rights to divert substantially more water than they currently use. Between 2000 and 2021, the average annual energy generation from the two major dams was 7.6 terawatt-hours (TWh)/year, enough to serve 2.5 million people. The river's landscapes and ecosystems provide critical habitat for federally protected species (*6*) and support an extensive recreation-based economy. Today, the entire flow is diverted along its 1400-mile course. In its lower reaches, only 10% of the natural flow reaches Mexico; rarely does the river flow to the Gulf of California (*7*).

CONSTRAINTS OF PAST POLICIES

Management of the river is governed by a set of interstate compacts, court decrees, federal laws, secretarial guidelines, and an international treaty that is collectively referred to as the Law of the River. The cornerstones are the 1922 Colorado River Compact and the 1944 Treaty between the United States and Mexico. The Compact is an agreement among seven Basin States, which divided the watershed into two parts, a lower basin that includes portions of Arizona, Nevada, and California and an upper basin that includes portions of Colorado, New Mexico, Utah, Wyoming, and a small area in Arizona. The Compact apportioned 7.5 million acre-feet (MAF; 1 acre-foot = 1233 m^3) per year of consumptive use to each basin and specified the division between them as Lees Ferry in northern Arizona (*8*, *9*). The Lower Basin was developing rapidly while Upper Basin development lagged; hence, this apportionment sought a degree of future equality among the basins. The Compact also required the Upper Basin not to deplete the river's flow to less than 75 MAF during any 10 consecutive years (the "non-depletion obligation") and required each basin to equally share any obligations to Mexico. The 1944 Treaty established a delivery requirement of at least 1.5 MAF/year.

The distinction between the Upper and Lower Basin created an institutional division that endures today. Lake Mead is often perceived as the water supply for the Lower Basin, and Lake Powell is primarily managed to avoid violation of the non-depletion obligation, even though all stored water effectively flows to the Lower Basin. This division is reinforced by Reclamation's institutional structure and distinct energy marketing arrangements between the two hydropower facilities.

Under the Law of the River, a total of 16.5 MAF/year of the mainstem flow is allocated for consumptive use. The primary metric used to evaluate hydrologic conditions is the natural flow at Lee Ferry. The Compact negotiators optimistically presumed a natural flow at Lee Ferry of 17.5 MAF/year and more than 20 MAF/year basin-wide. Evidence suggests, however, that they eschewed scientifically sound estimates that the available supply was potentially less (*9*). This knowledge was dismissed to help reach an agreement; the basin is increasingly paying the price for this strategy. The 20th-century natural flows at Lees Ferry averaged 15.2 MAF/year, an amount nearly sufficient to meet the Upper Basin's peak use of 4.0 MAF/year, 9.0 MAF/year of normal allocation in the Lower Basin and Mexico, plus 2.4 MAF/year for typical evaporation losses. However, since 2000, the average natural flow dropped to 12.3 MAF/year. To continue meeting demands, storage in lakes Powell and Mead decreased from 46 to 13.8 MAF. If the Millennium Drought continues or inflows decline further, then the only option will be to reduce consumptive uses to match the diminished supply.

THE RACE TO REDUCE DEMANDS

The Lower Basin and Mexico have been fully using their combined 9.0 MAF/year apportionment of the Colorado River. Under forceful federal prompting, the Lower Basin states committed in 2007 to reductions in consumptive uses (known as "shortages") in stages based on Lake Mead levels through "Interim Guidelines." Recognizing that these reductions would be insufficient to slow the drawdown of Lake Mead, a 2019 Drought Contingency Plan (DCP) augmented these commitments. Mexico agreed to reduce uses

[1]Environmental Change Institute, University of Oxford, Oxford, UK. [2]Water Balance Consulting, Boulder, CO, USA. [3]Colorado Water Center, Colorado State University, Fort Collins, CO, USA. [4]Utah Water Research Laboratory and Department of Civil and Environmental Engineering, Utah State University, Logan, UT, USA. [5]Independent consultant. [6]Center for Colorado River Studies, Department of Watershed Sciences, Utah State University, Logan, UT, USA. Email: kevin.wheeler@ouce.ox.ac.uk

approximately in proportion to US commitments through negotiated implementation agreements to the 1944 Treaty (Minutes).

Collectively, these agreements require the Lower Basin and Mexico to reduce their 9.0 MAF/year usage by 2.7 to 15.2% (0.241 to 1.375 MAF/year), with reductions increasing as Lake Mead's storage declines from 41 to 23% full (10.9 to 6.0 MAF). Reductions in Lower Basin use already occurred in 2020 and 2021 under the DCP. Additional voluntary reductions of 0.5 MAF/year by Lower Basin States and 0.1 MAF/year by Reclamation were recently proposed (*10*).

The Interim Guidelines and Treaty Minutes were triggered for the first time in 2022; thus, the combination of required and newly proposed voluntary Lower Basin and Mexico reductions will be 13.5% of their allocation (1.213 MAF/year). If Lake Mead storage declines to 6.0 MAF (23% of capacity), then the required and voluntary reductions would reach 21.9% (1.975 MAF/year). Citing concerns of hydropower failure, Reclamation's commissioner Touton informed Congress in June that 2 to 4 MAF of reductions below current commitments are needed. She did not specify how these reductions should be made among the states but reiterated the federal authority to act unilaterally if needed. All interstate and international shortage agreements will expire by 2026; a renegotiation process is underway.

THE UPPER BASIN SQUEEZE

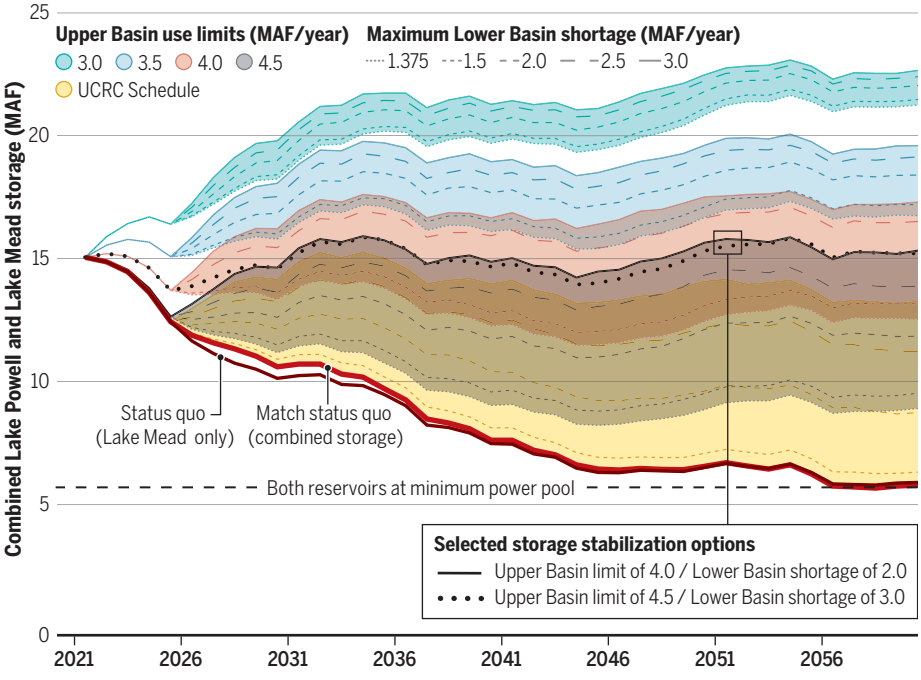
In contrast to the Lower Basin and Mexico, the Upper Basin is not using its full 7.5 MAF/year apportionment. Between 2000 and 2020, Upper Basin consumptive uses averaged 3.7 MAF/year plus at least 0.7 MAF/year of reservoir evaporation. There are plans for additional development; the Upper Colorado River Commission (UCRC) ambitiously projects 5.4 MAF/year of Upper Basin uses by 2060, exclusive of reservoir evaporation (*11*). Additional Upper Basin water use threatens to expose the uncertainty around the meaning of the Compact's non-depletion obligation, which in turn could upset basin-wide water delivery expectations.

Under variable year-to-year hydrologic conditions but with unchanging mean flows, the non-depletion obligation is frequently interpreted as a firm requirement for the Upper Basin to deliver a fixed volume downstream. Under declining flows, however, the meaning of a non-depletion obligation becomes unclear. A fixed delivery requirement under declining flows puts the entire burden of climate change on the Upper Basin. A more nuanced view of this obligation—and one that would arguably align with the Compact negotiators' intentions—is that a delivery obligation applies only to intermittent drought risk with no underlying change in mean flows, not the substantively different and much larger risk of permanently reduced flows.

One thing is clear: Additional Upper Basin consumptive uses would decrease inflows to Lake Powell and reduce storage volumes in lakes Powell and Mead. Lower Basin users have indicated that they are unlikely to reduce their uses to stabilize reservoirs only to see new upstream uses nullify these conservation efforts.

Average combined storage assuming drought conditions continue

Average end-of-year combined Lake Powell and Lake Mead storage is shown, assuming hydrologic conditions of the Millennium Drought continue. Results show combined reservoir contents using a range of Upper Basin consumptive use limits (colored ribbons) along with a range of Lower Basin maximum consumptive use reductions (line styles) triggered when the combined storage falls below 15 million acre-feet (MAF). The status quo lines use the 2016 Upper Colorado River Commission (UCRC) projections and existing elevation-based shortage triggers. All water use and shortage values are annual volumes (MAF/year).



WHAT MUST BE DONE?

Considering alarming warming trends and the past 23 years of drought, water managers must face the possibility that recent conditions will persist or worsen. Tree ring studies indicate that longer past droughts have occurred (*12*). Up to half of the recent flow decline has been attributed to Upper Basin warming, and additional declines are likely with continued climate trends (*1*, *2*). Under these conditions and with reservoirs nearly depleted, simple mass balance dictates that consumptive uses must be reduced. But to what extent and how should reductions be allocated?

The Upper Basin emphasizes water use equality between the basins as envisioned in the Compact, yet 100 years later, that has not occurred. Economic and equity considerations also exist. The Lower Basin irrigates less than half the area irrigated by the Upper Basin, yet its agricultural sales are more than three times that of the Upper Basin (*13*). Because the loss of an established resource is arguably more harmful than never having developed one, proposed large new uses are being questioned, and existing uses are facing unprecedented reductions.

Given many possible solutions, our research identified combinations of Upper Basin consumptive use limitations and Lower Basin reductions to maintain reservoir storage levels if the Millennium Drought continues (*14*). If these measures allow the current storage levels to be maintained, then we consider the system to be stabilized under these specific, but highly relevant, runoff conditions. Although the focus of our study is a scenario of continued drought, the insights and approaches can be adapted to plan for other future scenarios.

We used Reclamation's Colorado River Simulation System (CRSS) (*15*), which has been used for all major basin-wide analyses and decisions on the Colorado for the past 20 years and will be used in forthcoming

GRAPHIC: K. FRANKLIN/*SCIENCE*

renegotiations. To represent future hydrologic conditions, we developed 100 possible scenarios by randomly resampling natural flows that occurred from 2000 to 2018 (*12*). Mean annual flow at Lee Ferry during this period was 12.4 MAF/year, and our resampling method maintained the annual variability (see supplementary materials).

The first management strategy we assessed is what would happen if current consumptive use reduction commitments by the Lower Basin and Mexico remain in place and uses in the Upper Basin increase as projected by the UCRC. This status quo scenario assumes continued drought conditions but otherwise uses all existing assumptions and logic in CRSS, including current obligatory and voluntary consumptive use reduction measures. Following Reclamation's previous studies, this scenario also assumes that the Upper Basin non-depletion obligation in the 1922 Compact is not invoked. Meeting this obligation would require the Upper Basin to curtail uses, a legally disputed issue owing to differing interpretations of when this obligation is triggered and to what extent Upper Basin uses must be curtailed. This issue is unlikely to be resolved in a time frame conducive to managing further drought. The result of the status quo scenario is sharply declining combined storage of lakes Mead and Powell (see the figure), which falls to levels that further threaten hydropower production and increases risk of disruptions to downstream water deliveries.

To evaluate the magnitude of potential policy changes to stabilize the system, we conducted a two-dimensional sensitivity analysis that mapped out a range of greater reductions in existing consumptive use by the Lower Basin and Mexico when reservoir storage diminishes, combined with a range of future Upper Basin uses. Our proposal deviated from current management practice in which Lower Basin shortages are based only on elevations of Lake Mead, a policy that reflects the institutional divisions between the Upper and Lower Basin. We instead used the combined storage of the two reservoirs to trigger consumptive use reductions to the Lower Basin and Mexico. Our approach acknowledges the hydrologic reality that water stored in both reservoirs is consumed almost exclusively in the Lower Basin and Mexico. Furthermore, the current operational policies that govern the storage balance between the reservoirs are likely to evolve in the forthcoming negotiations. We also assumed that the non-depletion obligation is not invoked if the Upper Basin limits future depletions. This removes the longstanding ambiguity over the meaning of the non-depletion obligation in exchange for defined Upper Basin use limits, providing both basins with less risk and more certainty.

To implement our approach, different combined storage trigger thresholds were iterated with varying Lower Basin and Mexico shortage volumes until a reasonable match with the status quo was discovered (see the figure). We then incrementally increased the consumptive use reductions to the Lower Basin and Mexico whenever the combined reservoir storage falls below 15 MAF. Reclamation's well-established CRSS model is thoroughly documented (*15*), and our adaptations are described in preceding work (*14*) and the supplemental materials.

If the Millennium Drought persists, then the combined storage under the status quo will decrease to 6 MAF (12% of total Mead and Powell storage) before it stabilizes. At this volume, either Glen Canyon Dam or Hoover Dam would stop generating hydropower. These impacts show reservoir contents averaged across conditions since 2000; exceptionally dry years such as 2020 and 2021 will have an even greater impact.

Current reservoir storage levels could, however, be stabilized if consumptive uses decrease under different scenarios (see fig. S1). If the Upper Basin commits to limit water uses to 4.5 MAF/year (60% of their 7.5 MAF/year allocation, approximately 0.8 MAF/year higher than recent use), then the Lower Basin and Mexico must commit to more than doubling their current maximum reductions in existing use to 3.0 MAF/year (see the figure and fig. S1). In this scenario, the Lower Basin and Mexico receive 66.7% of their allocation, nearly matching the Upper Basin percentage. If the Upper Basin limits their depletions to 4.0 MAF/year (53.3% of their allocation, 0.3 MAF/year higher than recent use), then the Lower Basin and Mexico would need to decrease uses by approximately 2.0 MAF/year to stabilize the reservoirs (see the figure and fig. S1), assuring 77.8% of their allocation. This is close to recently proposed maximum Lower Basin and Mexico commitments to reduce existing use, which would not be invoked until Lake Mead declines further by 3 MAF. Delaying these reductions until then would result in greater loss of storage and stabilization occurring at lower levels than shown in the figure.

Water management models such as CRSS are only one part of the difficult work needed to achieve real-world solutions. Resolving complex water supply problems in large transboundary basins also requires deep understanding of the social and economic implications of any proposed policies, along with political barriers to adoption. Such work is iterative and slow, adding to the difficulties and pressures faced by decision makers.

Our results show that although current policies are inadequate to stabilize the Colorado River if the Millennium Drought continues, various consumptive use strategies can stabilize the system. However, these measures must be applied swiftly. Although these concessions by both basins may seem unthinkable at present, they will be necessary if recent conditions persist. ■

REFERENCES AND NOTES

1. B. Udall, J. Overpeck, *Water Resour. Res.* **53**, 2404 (2017).
2. P. C. D. Milly, K. A. Dunne, *Science* **367**, 1252 (2020).
3. S. Blumstein, J. D. Petersen-Perlman, *Water Int.* **46**, 306 (2021).
4. J. Fleck, A. Castle, *Water* **14**, 2 (2022).
5. US Bureau of Reclamation (USBR), "Colorado River 5-year probabilistic projections report" [US Department of the Interior (DOI), May 2022]; https://www.usbr.gov/lc/region/g4000/riverops/crss-5year-projections.html.
6. K. L. Dibble, C. B. Yackulic, T. A. Kennedy, K. R. Bestgen, J. C. Schmidt, *Ecol. Appl.* **31**, 1 (2021).
7. J. Pitt *et al.*, *Ecol. Eng.* **106**, 629 (2017).
8. Acre-feet is the volumetric unit historically embedded in all policies, legal allocations, and operational decisions on the Colorado River.
9. E. Kuhn, J. Fleck, *Science be Dammed: How Ignoring Inconvenient Science Drained the Colorado River* (Univ. Arizona Press, 2019).
10. USBR, Arizona Department of Water Resources, Southern Nevada Water Authority, Press Release: Water agencies announce partnership to invest $200 million in conservation efforts to bolster Colorado River's Lake Mead, under 500+ plan (Central Arizona Project, Phoenix, 2021); https://library.cap-az.com/documents/departments/planning/colorado-river-programs/CAP-500PlusPlan-NewsRelease.pdf.
11. Upper Colorado River Commission, "Upper Colorado River division states, current and future depletion demand schedule" (2016); http://www.ucrcommission.com/RepDoc/DepSchedules/CurFutDemandSchedule.pdf.
12. H. Salehabadi *et al.*, "The future hydrology of the Colorado River Basin" (Center for Colorado River Studies, Utah State University, 2020); doi:10.4211/hs.d3efcf0c930646fd9ef4f17c56436d20.
13. USBR, "Moving Forward Effort: Phase 1 Report" (DOI, May 2015); https://www.usbr.gov/lc/region/programs/crbstudy/MovingForward.
14. K. Wheeler *et al.*, "Alternative management paradigms for the future of the Colorado and Green Rivers" (Center for Colorado River Studies, Utah State University, 2021); doi:10.4211/hs.59175cf99a58462f901cdf56ec79ddbe.
15. USBR, "DRAFT CRSS: Key modeling assumptions, June 2021 model" (DOI, 2021); http://bor.colorado.edu/Public_web/CRSTMWG/CRSS.

ACKNOWLEDGMENTS

We acknowledge the valuable support of A. Castle, D. Kenney, and D. Tarboton. The authors acknowledge funding from the Catena Foundation (grant 202059), Walton Family Foundation (grant 2018-585), My Good Fund, and David Bonderman fund. J.C.S. acknowledges a Janet Quinney Lawson Chair in Colorado River Studies endowment. H.S. acknowledges funding from the Utah Water Research Laboratory. K.G.W. acknowledges funding from the Oxford Martin Programme on Transboundary Resource Management.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo4452

10.1126/science.abo4452

BOOKS *et al.*

Adding structure can offer students more opportunities to engage, practice, and get feedback.

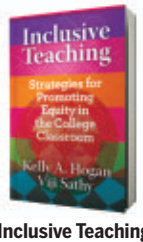
EDUCATION

Setting college students up for success

A pair of researchers outline strategies for ensuring that postsecondary courses are inclusive

By **Jeremy L. Hsu**

What does it mean to have an inclusive college classroom? How can we promote inclusion—both inside and outside the classroom—and help students from different backgrounds succeed in science, technology, engineering, and mathematics (STEM)? Authors Kelly Hogan and Viji Sathy, both STEM professors at the University of North Carolina at Chapel Hill and well-known education researchers, tackle these critical questions in their aptly named book, *Inclusive Teaching*.



Inclusive Teaching
Kelly A. Hogan and Viji Sathy
West Virginia University Press, 2022. 272 pp.

Hogan and Sathy cite the large disparities in performance often observed across different student demographics in undergraduate STEM courses as inspiration for their work, regularly returning to these disparities to convince readers that more must be done to ensure that all students are equipped to succeed. The book provides evidence-based practices and actionable suggestions on inclusive teaching, offering a timely and much-needed complement to the existing literature.

The authors shine at translating the sometimes-dense literature of education research into clear, accessible, and actionable steps for instructors. In chapter 3, for example, Hogan and Sathy focus on language used in course syllabi, an element that is oft forgotten. Here, they cite studies showing how student-centered language (e.g., language that uses "I," "you," and "we" rather than "students will") can make students feel more connected and perceive the instructor as more competent. They also provide specific examples and rubrics for improving syllabi. Each chapter closes with a concise summary, framed as a checklist of items for instructors to consider.

Hogan and Sathy weave first-hand accounts into the narrative, describing the continual improvements they have made in their teaching over their respective careers. For example, in the first chapter, Hogan describes adding guided reading questions, pre-class homework, and in-class polls to her introductory biology course after seeing certain students struggle. Sathy, meanwhile, describes how student complaints about her class's pace—some saying it proceeded too quickly, some saying too slowly—inspired her to create videos of herself solving problems that students could watch at their own pace before class.

Hogan and Sathy also include summaries of education research—their own as well as work conducted by others—and fictional scenarios that highlight the impact of inclusive practices. This juxtaposition of educational research findings with hypothetical scenarios is jarring at first. However, as the book progresses, this duality proves effective, allowing the authors to discuss empirical evidence of the impacts of inclusive teaching practices while clearly illustrating how these impacts can play out in the real world. In chapter 2, for example, the authors discuss how structure can improve learning outcomes, by providing more required opportunities for students to engage, practice, and get feedback. They illustrate this potentially abstract concept with a fictional professor, Dr. Slim, who offers limited chances for students to practice and get feedback in the course. They argue that by adding more structure in the form of pre-class questions, online quizzes, group discussion, in-class activities, and post-class reflections, Dr. Slim can guide and support students more effectively.

At times, however, this strategy is less convincing, particularly when the authors discuss areas with less empirical work from which to draw. There is a paucity of research examining instructor practices outside the classroom, for example. The authors rely on their own experiences to fill such gaps, although the distinction between anecdote and evidence could have been made clearer.

The book addresses several underexplored, but critical, areas where instructors can play vital roles in promoting inclusion outside the classroom. A whole chapter, for instance, focuses on modeling inclusivity with students, including suggestions for proactively emailing struggling students; promoting transparency and sharing norms about office hours; reducing bias in grading; and providing structured feedback to encourage students. Similarly, another chapter is devoted to institutional change, discussing how instructors can reflect and document their inclusive teaching practices and use their efforts to advocate for change.

These strengths make *Inclusive Teaching* compelling and critical. Given the urgent need to promote justice, equity, diversity, and inclusion in our communities, the book is a must-read for all who are in a position to better support inclusive teaching inside and outside the classroom. ■

10.1126/science.abq5892

The reviewer is at the Schmid College of Science and Technology, Chapman University, Orange, CA 92866, USA. Email: hsu@chapman.edu

The virtual worlds of the metaverse

An immersive internet is just around the corner, for better or worse

By **Dov Greenbaum**

With the cryptocurrency market spiraling downward and the bursting of the non-fungible token (NFT) bubble, many skeptics believe that the metaverse—a thoroughly immersive emerging version of the internet—is just another fleeting trend. Mathew Ball's *The Metaverse* seeks to reassure the reader that it is not. Reporting that the term "metaverse" was mentioned more than 260 times in US Securities and Exchange Commission filings in 2021, Ball suggests that "the sheer number of companies that see potential value in the Metaverse speaks to the size and diversity of the opportunity."

The three-part book provides more than just a definitive definition of the metaverse, which, despite mounting interest, many people still fail to fully understand. The first section delivers an overview of the technology in all its potential iterations. The second examines the ongoing technological infrastructure expansion that is necessary to grow and maintain it. And the final section predicts the societal changes that will arise as a result of expansive adoption of the metaverse.

Ball defines the metaverse as "a massively scaled and interoperable network of real-time rendered 3D virtual worlds that can be experienced synchronously and persistently by an effectively unlimited number of users with an individual sense of presence, and with continuity of data." He refers back to this definition throughout the book to signpost various central aspects and elements of the technology as he details its potential. Ball also frequently and deferentially references the 1992 novel *Snow Crash* and its author, Neal Stephenson, who coined the term metaverse, as he presents the technology's evolution through the lens of the development of the internet and the online video gaming industry.

In contrast to the open nature of the internet, the metaverse will be composed, at least initially, of many closed and proprietary systems, resulting in competing virtual worlds. It will likely exacerbate many of the ongoing ills of the internet, including perpetuating a lack of data rights and facilitating manipulation, radicalization, harassment, and misinformation online. But according to Ball, with the necessary governance, the benefits will ultimately outweigh the system's shortcomings.

There are also considerable technical hurdles that require resolution before the metaverse—at least as it is currently conceived of in popular culture—can be achieved. These include "fitting a supercomputer into the frame of…glasses," which Mark Zuckerberg has described as "the hardest technology challenge of our time." Ball personally considers the current inability to host more than a limited number of individuals concurrently in one version of a virtual world to be the hardest problem to solve.

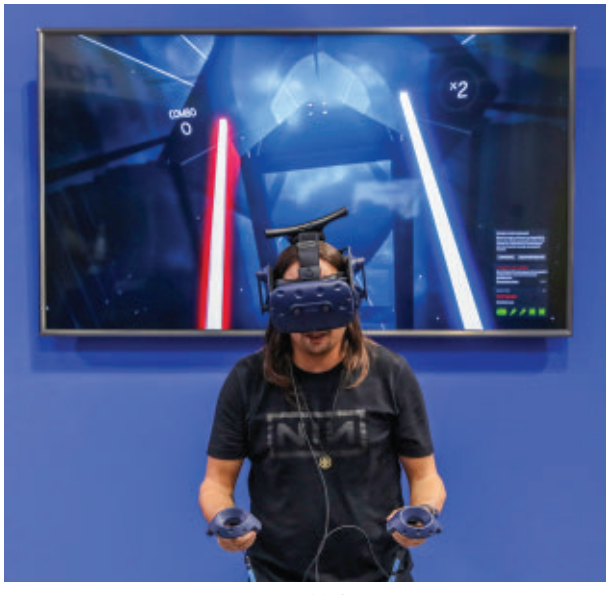
While much of the book is a breezy read, Ball sometimes wades into the details of what systems will be necessary for the metaverse to thrive. He describes, for example, the genesis and current state of internet payment rails, the "complex series of systems and standards, deployed across a wide network and in support of trillions of dollars in economic activity." Here, he argues that the current control exerted by the major tech companies over payment systems will be the hardest nontechnical challenge for the metaverse to overcome. And while he seems agnostic as to the overall future of distributed ledger technology, he suggests that blockchains could become a key platform in reducing this control by the tech companies.

There is some debate as to when exactly the critical mass of relevant innovations will solve the technical and logistic barriers now precluding the metaverse. Like the internet before it, "Disruption is not a linear process, but a recursive and unpredictable one." But when it does arise, the book suggests that many may see it simply as a successor to the current internet.

Ball distinguishes the metaverse from another proposed internet successor—Web3—"a somewhat vaguely defined future version of the internet built around independent developers and users." The often-conflated concept aims to use blockchain technologies to become a more decentralized internet. Ball notes that while "the *principles* of Web3 are likely critical to establishing a thriving Metaverse," confusing the metaverse for Web3 is like "conflating the rise of democratic republics with industrialization or electrification"—one is about governance, the other is about technology.

The book also outlines three nontechnical, but nonetheless critical, factors that will be essential to metaverse growth. The first is regulatory action to open up better payment solutions and increase standardization and interoperability. Online games—particularly *Fortnite*, *Roblox*, *Minecraft*, and even *Microsoft Flight Simulator*—have been described as "proto-Metaverses," and they continue to provide a second essential nontechnical factor: social acceptance. The third and final component crucial to the success of the metaverse will be the availability of useful and desirable experiences, including those related to education, lifestyle, entertainment, fashion, advertising, and industrial innovation. ■

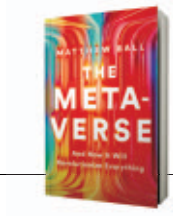
10.1126/science.add5905



A visitor plays a virtual game at the 2018 IFA consumer electronics fair.

PHOTO: SHAN YUQI/XINHUA/ALAMY

The Metaverse: And How It Will Revolutionize Everything
Matthew Ball
Liveright, 2022. 352 pp.



The reviewer is at the Zvi Meitar Institute for Legal Implications of Emerging Technologies, Harry Radzyner Law School, Reichman University, Herzliya, Israel, and Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA. Email: dov.greenbaum@aya.yale.edu

Ninth International Congress on Peer Review and Scientific Publication

September 8-10, 2022 | Chicago, IL

Register now!

Congress Advisory Board

The Ninth International Congress on Peer Review and Scientific Publication will be held September 8-10, 2022, in Chicago, Illinois, USA, in person and online. This congress, organized by JAMA Network, *The BMJ*, and METRICS, will feature 3 days of presentations of new and original research.

What?

The Congress will provide a forum for the presentation and discussion of new research into the quality and credibility of peer review and scientific publication, to establish the evidence base on which scientists can improve the conduct, reporting, and dissemination of scientific research.

Who?

Participants will include editors and publishers of scientific peer-reviewed journals, researchers, funders, bibliometric and informatics experts, information innovators, librarians, journalists, policymakers, ethicists, scientific information disseminators, and anyone interested in the progress of the scientific information enterprise and the quality of scientific evidence.

The Congress embraces a wide range of disciplines, including biomedicine, health and life sciences, applied sciences, basic sciences, physical and chemical sciences, mathematics, computer sciences, engineering, economics, and social sciences. New and emerging disciplines are also welcome.

When?

September 8-10, 2022

Registration will open to attendees in Spring, 2022.

Questions? jama-peer@jamanetwork.org

JAMA Network™ thebmj METRICS META-RESEARCH INNOVATION CENTER AT STANFORD

peerreviewcongress.org

A wildfire rages in the forests of the Zagros Mountains in southwestern Iran.

PHOTO: AHMAD HALABISAZ/XINHUA/GETTY IMAGES

Edited by **Jennifer Sills**

Better preparation for Iran's forest fires

In Iran, one of the driest countries in the world (*1*–*3*), about 1500 wildfires are reported each year, destroying thousands of hectares of forests and pastures annually and causing more than US$5.6 million in economic damage (*4*–*6*). To mitigate fire damage, Iran should improve containment strategies and work toward more effective fire prevention.

By the time fires are detected in Iran, they are often difficult to control (*5*, *7*). Iran lacks specialized manpower, forest emergency bases, and air relief. Firefighters often lack access to the fire sites and water reservoirs. As a result, time-consuming preparation hinders their ability to prevent the fire's spread (*3*, *7*, *8*).

Iran could also benefit from better fire prevention (*9*). By creating a database of past fires in the geographic information system, the Iranian government could focus preventive measures in the areas most at risk of wildfire (*10*). A spatial database could track the factors influencing the occurrence of wildfires and the environmental characteristics of forests in different regions. These data would allow Iran to build fire towers and watchtowers in high-risk areas. In addition, Iran should educate the public to raise awareness about forest vulnerability and the importance of natural resources (*3*, *5*).

Iran should invest the funds necessary to provide updated firefighting training and equipment. The government should also build forest road networks, water storage ponds, and helicopter launch pads throughout high-risk areas. Finally, nationwide wireless networks would facilitate fire alerts (*3*). With the support of the public, nongovernmental organizations, and scientists, Iran can reduce the rate of wildfires.

Mahmood Tavakoli Hafshejani,[1] Mahmoud Nasrollahzadeh[2], Valiollah Mirkhani[1]*
[1]Department of Chemistry, Faculty of Science, University of Isfahan, Isfahan 81746-73441, Iran.
[2]Department of Chemistry, Faculty of Science, University of Qom, Qom, Iran.
*Corresponding author.
Email: mirkhani@sci.ui.ac.ir

REFERENCES AND NOTES

1. A. Shahabfar, A. Ghulam, J. Eitzinger, *Int. J. Appl. Earth Obs. Geoinf.* **18**, 119 (2012).
2. R. Akbari, M. Nasrollahzadeh, *Science* **375**, 984 (2022).
3. H. Ghazanfarpoor, S. Hasanzadeh, M. Hamedi, *J. Nat. Environ. Hazards* **5**, 61 (2017) [in Farsi].
4. R. Abedi, *Int. J. Geoherit. Parks* **10**, 84 (2022).
5. S. Eskandari, S. Eskandari, *Hum. Environ.* **19**, 175 (2021) [in Farsi].
6. R. Jahdi *et al.*, *Nat. Hazards* **101**, 911 (2020).
7. S. Aleemahmoodi Sarab, J. Feghhi, H. Goshtasb, *Intl. J. Mol. Evol. Biodivers.* **3**, 15 (2013).
8. S. Ozmehmet Tasan, Y. E. Ergenc, in *New Global Perspectives on Industrial Engineering and Management*, J. Mula, R. Barbastefano, M. Díaz-Madroñero, R. Poler, Eds. (Springer, 2019), pp. 203–211.
9. A. Alexandridis, D. Vakalis, C.I. Siettos, G.V. Bafas, *Appl. Math. Comput.* **204**, 191 (2008).
10. Z. X. Zhang, H. Y. Zhang, D. W. Zhou, *J. Arid Environ.* **74**, 386 (2010).

10.1126/science.add5194

China's restoration fees require transparency

China is home to 10% of the world's wetland areas, but many of those wetlands are threatened by development (*1*). To increase conservation efforts, China's first wetland protection law, which came into force on 1 June (*2*), will charge a fee to developers whose projects result in wetland area losses. The fees will pay for restoring wetlands with comparable qualities and quantities elsewhere. When this strategy has been implemented in the past, transparency has been insufficient. The wetlands law, as well as other laws requiring restoration fees, must include data tracking and availability to ensure that the money is used as intended and that the restored ecosystems are suitable substitutes for those that have been degraded.

China has enacted two previous nationwide mandatory natural habitat restoration fees, one in 1998 for forest vegetation (*3*, *4*) and one in 2003 for grassland vegetation (*5*). In each case, tracking conservation outcomes and evaluating whether ecological compensation requirements and targets are being met have proved challenging. Information on how much money various levels of governments have collected and spent, and on what, is extremely limited, at least in the public domain. The lack of financial transparency could lead to misuse or misappropriation of restoration funds as well as ineffective use of funds, with money going toward, for example, projects with no evidence of positive outcomes (*6*).

A similar approach has been implemented for wetlands in the United States since the 1980s (*7*) as well as for other habitats in other countries, including Australia, Brazil, the United Kingdom, and Germany (*6*, *8*). In each case, results were mixed (*8*, *9*). Given that success is not guaranteed, it is even more vital to track the progress of the program and adjust its implementation to maximize benefits.

In 2021, China committed to enhancing biodiversity and ecosystem functions and services by gradually advancing information disclosure and encouraging public participation (*10*). In light of this pledge, China's government should create a mechanism to clearly, thoroughly, and regularly report the collection and use of forest, grassland, and wetland restoration fees. The information should include government spending, ecological assessment before development begins, restoration implementation, and outcomes (*11*), and all data should be made available for public scrutiny. As a monitoring system model, China could use the US Regulatory In-lieu Fee and Bank Information Tracking System, a registry of conservation-related programs that has been in place for nearly 40 years (*12*).

Shuo Gao[1,2]*, Joseph W. Bull[3], Zhiqiang Wu[4,5], Renlu Qiao[4,5], Li Xia[6], Ming K. Lim[7]
[1]Interdisciplinary Centre for Conservation Science, University of Oxford, Oxford, Oxfordshire, UK. [2]St. Hilda's College, University of Oxford, Oxford, Oxfordshire, UK. [3]Durrell Institute of Conservation and Ecology, University of Kent, Canterbury, Kent, UK. [4]College of Architecture and Urban Planning, Tongji University, Shanghai, China. [5]Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, Shanghai, China. [6]School of Management, University of Science and Technology of China, Hefei, Anhui, China. [7]Adam Smith Business School, University of Glasgow, Glasgow, UK.
*Corresponding author.
Email: shuo.gao@st-hildas.ox.ac.uk

REFERENCES AND NOTES

1. W. Xu *et al.*, *Curr. Biol.* **29**, 3065 (2019).
2. Ministry of Ecology and Environment of the People's Republic of China, "Wetland Protection Law of the People's Republic of China" (2022); www.mee.gov.cn/ywgz/fgbz/fl/202112/t20211227_965347.shtml [in Chinese].
3. Ministry of Ecology and Environment, "Forestry Law of the People's Republic of China" (2021); www.mee.gov.cn/ywgz/fgbz/fl/202106/t20210608_836755.shtml [in Chinese].
4. B. Madsen, N. Carroll, K. Moore Brands, "State of biodiversity markets report: Offset and compensation programs worldwide" (Ecosystem Marketplace, Washington, DC, 2010).
5. Ministry of Ecology and Environment, "Grassland Law of the People's Republic of China" (2021); www.mee.gov.cn/ywgz/fgbz/fl/200212/t20021228_81958.shtml [in Chinese].
6. J. W. Bull, N. Strange, *Nat. Sustain.* **1**, 790 (2018).
7. R. F. Ambrose, *Wetlands Australia J.* **19**, 1 (2010).
8. S. O. zu Ermgassen *et al.*, *Conserv. Lett.* **12**, e12664 (2019).
9. W. Matthews, A. G. Endress, *Environ. Manage.* **41**, 130 (2008).
10. State Council of the People's Republic of China (State Council), "Opinions on Further Strengthening Biodiversity Conservation" (2021); www.gov.cn/zhengce/2021-10/19/content_5643674.htm [in Chinese].
11. H. Kujala *et al.*, *One Earth* **5**, 650 (2022).
12. U.S. Army Corps of Engineers, U.S. Environmental Protection Agency, U.S. Fish and Wildlife Service, National Marine Fisheries Service, Natural Resources Conservation Service, "Federal guidance for the establishment, use and operation of mitigation banks," *Fed. Reg.* **60,** 58605 (1995).

10.1126/science.add5125

Global goals overlook freshwater conservation

As global conservation and restoration policies focus on a land and sea framework, freshwater biodiversity and services continue to decline at alarming rates (*1*). If freshwater ecosystems are overlooked, their sustainability could be compromised when decision-makers evaluate trade-offs with land and sea conservation and development goals. To protect freshwater biodiversity and vital services, international agreements must explicitly acknowledge freshwater ecosystems as a unique realm and set specific goals to address their problems (*2*, *3*).

At the 2021 UN Climate Change Conference in Glasgow (COP26), countries reaffirmed their commitments to the three Rio Conventions on Biological Diversity, Climate Change, and Desertification (*4*). The three respective panels are preparing reports that will shape the 2030 sustainable development goals (SDGs) and the 2021–2030 UN Decade on Ecosystem Restoration. Setting explicit objectives for freshwater ecosystems in these goals must be a priority.

Unfortunately, the recently released "Global land outlook" (*5*), the flagship publication of the UN Convention to Combat Desertification, a convention that defines pathways to sustainable land and water management, still mostly treats fresh water as a simple resource for services such as irrigation and consumption rather than a unique ecosystem that sustains biodiversity and a range of other services and that has particular management needs. The undervaluing of freshwater ecosystems is demonstrated by how rarely the word is used: Fresh water is mentioned twice in the summary for decision-makers, but both times as "freshwater use," with no mention of the associated ecosystems or their management. Land restoration commitments of "1 billion hectares of farms, forests, and pastures" make no explicit allusion to rivers or other freshwater ecosystems. This shortsightedness is consistent with SDG 15 ("life on land"), which discounts the uniqueness of the freshwater realm, and with SDG 6 ("water and sanitation"), which prioritizes only the most immediate services that freshwater ecosystems provide. Underestimating the value of fresh water undermines the potential for long-term sustainability.

Some recent reports provide hope that we can prioritize freshwater conservation and recognize the unique problems and challenges that such ecosystems face. In the Intergovernmental Panel on Climate Change's sixth assessment report (AR6), the working group on "impacts, adaptation and vulnerability" breaks ecosystem impacts into terrestrial, ocean, and fresh water (*6*). In addition, the latest draft of the post-2020 Global Biodiversity Framework indicates the possibility of including fresh water in several goals and targets (*7*).

Ahead of keystone events like COP27 in November in Sharm El-Sheikh, Egypt and the UN Biodiversity Conference (COP15) finally scheduled for December in Montreal, authors of reports that influence international agreements must make the case that freshwater ecosystems require attention independent of other conservation efforts. This recognition could include, if not an additional SDG, targets addressing freshwater-specific area protection and restoration, the waterflow quality needed to maintain ecosystems and related services, and integrated water resources management (*2*).

Ground and surface freshwater habitats are home to more than 10% of all known species, including 30% of all vertebrates (*8*). The ecosystem services they provide are estimated to be worth more than US$4 trillion annually (*9*). Only by explicitly recognizing the value and distinctiveness of freshwater ecosystems can we set goals that can effectively protect them.

Duarte V. Gonçalves[1,2]* and Virgilio Hermoso[3]
[1]Interdisciplinary Centre of Marine and Environmental Research (CIIMAR), Matosinhos, Portugal. [2]Research Centre in Biodiversity and Genetic Resources (CIBIO/InBIO), University of Porto, Vairão, Portugal. [3]Departamento de Biología Vegetal y Ecología, Universidad de Sevilla, Sevilla, Spain.
*Corresponding author.
Email: duartenvg@gmail.com

REFERENCES AND NOTES

1. World Wildlife Fund, "Living planet report 2020: Bending the curve of biodiversity loss" (Gland, Switzerland, 2020).
2. C. B. van Rees *et al.*, *Conserv. Lett.* **14**, e12771 (2021).
3. Wetlands-International, "We need wetlands: The urgent case for global wetland targets" (2022).
4. "Glasgow leaders' declaration on forests and land use," UN Climate Change Conference UK 2021 (2021); https://ukcop26.org/glasgow-leaders-declaration-on-forests-and-land-use/.
5. UN Convention to Combat Desertification (UNCCD), "The Global Land Outlook, second edition" (UNCCD, Bonn, 2022).
6. Intergovernmental Panel on Climate Change, "Climate change 2022: Impacts, adaptation, and vulnerability. Contribution of Working Group II to the sixth assessment report of the Intergovernmental Panel on Climate Change" (Cambridge University Press, 2022).
7. Convention on Biological Diversity, "Preparation of the post-2020 global biodiversity framework" (2022); www.cbd.int/doc/c/c949/b2cc/a311c0c411d3a81134e2c7f3/wg2020-03-l-02-en.pdf.
8. R. A. Mittermeier *et al.*, "Fresh water: The essence of life (vol. 18)" (Conservation International, 2010).
9. A. J. Lynch *et al.*, *Environ. Rev.* **24**, 115 (2016).

10.1126/science.add6295

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

PHOTO: MICHAEL DURHAM/MINDEN PICTURES

EVOLUTION

Maintaining difference

Species often comprise several ecotypes, distinct populations that occupy different habitats. Ecotypes can persist over long time periods, even with substantial gene flow between them, which raises the question of how they maintain their locally adaptive phenotypes over time. Hager *et al*. examined the genetic basis of two traits, tail length and coat color, that define the forest and prairie ecotypes of deer mice. They found a large chromosomal inversion that links redder coats and longer tails in the forest ecotype. Modeling suggests that the inversion originated under divergent selection many thousands of generations ago and likely provided a benefit to the forest ecotype by suppressing recombination despite gene flow. —BEL *Science*, abg0718, this issue p. 399

Deer mice, such as this individual from central Oregon, are a widely dispersed species with locally adaptive ecotypes.

SURFACE CHEMISTRY

Making surface chemistry more exact

Accurate description of elementary steps of chemical reactions at surfaces is a long-standing challenge because of the lack of reliable experimental measurements of the corresponding rate constants, which also makes it impossible to rigorously validate theoretical estimates. Even for reactions as simple as thermal recombination of hydrogen atoms on platinum surfaces, previous experimental rate constants have only been obtained with large uncertainties. Using velocity-resolved kinetics and ion imaging–based calibration of absolute molecular beam fluxes, Borodin *et al*. managed to overcome established experimental difficulties and report unprecedentedly accurate rate constants for this reaction over a wide temperature range. They also demonstrate a parameter-free model that quantitatively reproduces the experiment, opening up new vistas for the growing field of computational heterogeneous catalysis. —YS *Science*, abq1414, this issue p. 394

OPTICS

Amplification in photonic time crystals

Regular photonic crystals are structures in which the refractive index is spatially periodic and can suppress the spontaneous emission of light from an emitter embedded in the structure. In photonic time crystals, the refractive index is periodically modulated in time on ultrafast time scales. Lyubarov *et al*. explored theoretically what happens when an emitter is placed in such a time crystal (see the Perspective by Faccio and Wright). In contrast to the regular photonic crystals, the authors found that time crystals should amplify emission, leading to lasing. —ISO *Science*, abo3324, this issue p. 425; see also abq5012, p. 368

ORGANIC CHEMISTRY

Tetrodotoxin by cycloaddition

Tetrodotoxin is a potent bacterial neurotoxin widely associated with pufferfish and thoroughly studied for its sodium channel–blocking properties. Its intricate structure of oxygen-rich interconnected rings has also long intrigued synthetic chemists. Konrad *et al*. report a comparatively concise route to the natural product from a glucose derivative. A dipolar cycloaddition enabled the formation of the cyclohexyl core at a later stage than prior approaches. Ruthenium catalysis was then key in assembling the surrounding oxygenated rings. —JSY *Science*, abn0571, this issue p. 411

SEMICONDUCTORS

Swift carriers

Boron arsenide is a semiconductor with several interesting properties, including a high thermal conductivity.

Theoretical calculations also suggest that it has high ambipolar mobility, a measure of the mobility of electrons and holes. Yue *et al*. and Shin *et al*. used different types of measurements to observe a high ambipolar mobility in very pure cubic boron arsenide. Shin *et al.* were able to simultaneously measure the high thermal and electrical transport properties in the same place in their samples. Yue *et al*. found even higher ambipolar mobility than the theoretical estimates at a few locations. Boron arsenide's combination of transport properties could make it an attractive semiconductor for various applications. —BG

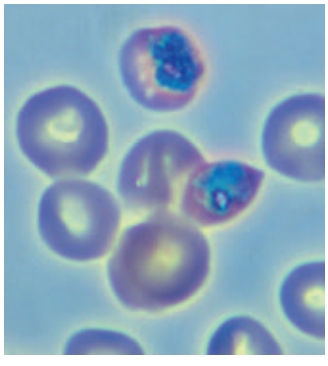
Science, abn4290 and abn4727, this issue p. 433 and p. 437.

MALARIA

Severe malaria insights

Diagnosis of children with severe malaria caused by infection with *Plasmodium falciparum* has been difficult in high-transmission settings because of the high coincidence of malaria with other febrile illnesses. Watson *et al*. analyzed data from 2649 severely ill children and adults in low- and high-transmission settings enrolled in several malaria clinical studies. Using a combination of platelet counts and plasma concentrations of *P. falciparum* histidine-rich protein-2 in a Bayesian latent class model, the authors achieved a sensitivity of approximately 74% and a specificity of approximately 93% in identifying severe malaria. Their findings revealed that one-third of children in high-transmission settings had severe febrile illness caused by other types of pathogens. —CNF

Sci. Transl. Med. **14**, eabn5040 (2022).



Microscopy image of *Plasmodium falciparium* parasites infecting red blood cells

TUMOR IMMUNOLOGY

Under the skin

Although systemic immune checkpoint blockade displays therapeutic efficacy in cancers, it induces immune-related adverse events that can undermine treatment. Van Pul *et al*. performed a phase 1 clinical trial testing the effects of an intradermal injection of antibodies that block cytotoxic T lymphocyte–associated protein-4 at the site of primary tumor excision in patients with early-stage melanoma. Seven of 13 patients immunologically responded to the treatment without severe adverse events. Responders had more tumor antigen–specific T cells in the blood, increased migratory dendritic cell activation in the sentinel lymph node, and decreased regulatory T cells in both the sentinal lymph node and the blood. This treatment has promise for patients with early-stage melanoma. —DAE

Sci. Immunol. **7**, eabn8097 (2022).

PROTEIN DESIGN

Designing around function

Protein design has had success in finding sequences that fold into a desired conformation, but designing functional proteins remains challenging. Wang *et al*. describe two deep-learning methods to design proteins that contain prespecified functional sites. In the first, they found sequences predicted to fold into stable structures that contain the functional site. In the second, they retrained a structure prediction network to recover the sequence and full structure of a protein given only the functional site. The authors demonstrate their methods by designing proteins containing a variety of functional motifs. —VV

Science, abn2100, this issue p. 387

IN OTHER JOURNALS

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

PLASTIC RECYCLING

Unlocking polystyrene

Many of today's most common plastics are difficult to break down for recycling. In polystyrene, for instance, the reactive C=C double bond in the monomer becomes an inert C–C single bond in the polymer. Kiel *et al.* sought to circumvent this problem by introducing a small percentage of alternative monomers into the backbone. Specifically, when styrene was copolymerized with 2.5% of a thionolactone, treatment of the plastic with a thiol cleaved the backbone at the sulfur sites. The resultant oligomers could then be repolymerized to show properties similar to those of pure polystyrene. —JSY

J. Am. Chem. Soc. 10.1021/jacs.2c05374 (2022).

Introducing a small percentage of certain alternative monomers into the backbone of polystyrene, such as the beads pictured, could make it easier to recycle.

CANCER

Targeting the tumor microenvironment

Cancer-associated fibroblasts (CAFs) are stromal cells with important roles in modulating tumorigenesis, tumor progression, and responses to therapy. CAFs produce collagen, a component of the extracellular matrix that can promote some types of cancer. Kay *et al*. examined CAFs from a human model and a mouse model of breast cancer and found that glutamine metabolism drives proline synthesis, which is required to produce tumorigenic collagens. The pyrroline-5-carboxylate reductase 1 (PYCR1) enzyme, important in proline synthesis, was overexpressed in CAFs. Reducing its expression

PHOTOS (LEFT TO RIGHT): MICHAEL ABBEY/SCIENCE SOURCE; TANAPHONGPICT/SHUTTERSTOCK

reduced collagen production, tumor growth, and progression in vivo. Numerous subtypes of CAFs with different functions exist in the tumor microenvironment, so targeting their pro-tumorigenic effects may be an effective strategy in certain cancer types. —GKA

Nat. Med. **4**, 693 (2022).

AGING

Fat to brain longevity signal

Lysosomes, once mainly thought to degrade cellular components, are now implicated in complex signaling roles. Savini *et al.* found that overexpression of the lysosomal acid lipase LIPL-4 (which serves as a fat storage function) in the intestine of the worm *Caenorhabditis elegans* seems to extend life span through communication with the brain. In the proposed scheme, a specific polyunsaturated fatty acid (PUFA) produced by LIPL-4 appears to be carried to the brain by the fatty acid–binding protein LBP-3. In the brain, the PUFA seemed to enhance transcription of the neuropeptide NLP-11, which is required for the life-extending effect. Thus, fatty acids bound to carrier proteins might have a hormone-like function, so lysosomes may be a signaling hub coordinating communication between organs. —LBR

Nat. Cell Biol. **24**, 906 (2022).

MALARIA

Digital gene deletion diagnosis

Rapid diagnostic tests for falciparum malaria primarily rely on detecting the virulence-related histidine-rich proteins 2 and 3 (HRP2 and HRP3). Sometimes, the genes *hrp2 and hrp3* are deleted in the parasite and the diagnosis is missed, which obviously presents a problem for treatment, surveillance, and control. Vera-Arias *et al.* have developed a high-throughput droplet digital polymerase chain reaction assay to detect *hrp2/3* deletions. This droplet technology is beginning to be adopted for malaria diagnosis in endemic countries. Screening showed that samples obtained from South America and Africa are highly heterogeneous for *hrp* deletions. For example, the authors found that up to 80% of parasites in hospitalized patients in Eritrea had *hrp2* deletion, and high levels of deletions also occurred in Brazil. Apart from avoiding false negatives, additional advantages of this approach lie in improved *Plasmodium falciparum* identification and a reduction in the number of assays needed for accurate diagnosis. —CA

eLife **11**, e72083 (2022).

CANCER

p53 in elephants

As we age, our cells are more likely to become cancerous because they accumulate more mutations. How very large, long-lived animals such as elephants and whales protect themselves against cancer is an interesting question. One hypothesis is that these animals have more copies or higher expression of the p53 tumor suppressor transcription factor. Indeed, elephants have as many as 20 copies of p53 genes, more than any other animal. Therefore, Padariya *et al.* combined in silico modeling and in vitro assays to explore the p53 isoforms of elephants. The authors discovered that they are functionally diverse and express a variety of binding motifs known as BOX-I MDM2. These motifs enhance sensitivity to cellular damage that prefigure cancer and promote defense pathways. —DJ

Mol. Biol. Evol. 10.1093/molbev/msac149 (2022).

GLOBAL WARMING

What are the odds?

How likely is it that we will be able to limit global warming to 1.5° or 2.0° C above the preindustrial average, the goals set by the Paris Agreement? Dvorak *et al.* used an emissions-based climate model to estimate the evolution of global temperature after an abrupt cessation of anthropogenic greenhouse gas emissions using a variety of emission scenarios. They found that there is only one chance in three that peak global warming can be kept below 1.5°C by 2027–2032 under all emissions scenarios, lending further urgency to efforts to eliminate our greenhouse gas contributions. —HJS

Nat. Clim. Change **12**, 547 (2022).

SENSORS

Something smells fishy here

Rotting fish release a bouquet of chemicals, but trimethylamine (TMA) is the main source of the odor that we detect. Sensors for TMA are thus used to indicate freshness but struggle to remain flexible and sensitive at low temperatures. Li *et al.* developed a sensor based on the MXene $Ti_3C_2T_x$, where T is a terminating group consisting of O, OH, or F. The MXene was modified with Au nanoparticles and then polymerized with a hydrogel and soaked in a solution of ethylene glycol and HCl to endow it with antifreeze properties. The composite hydrogel showed high strength, stretchability, and toughness even at low temperatures, and was able to sense TMA at both room temperature and 0°C. —MSL

ACS Appl. Mater. Interfaces **14**, 30182 (2022).

PHOTO: YASUYOSHI CHIBA/AFP/GETTY IMAGES

Long-lived animals such as elephants (shown here is Tolstoy, a 49-year old "tusker" in Amboseli National Park) thwart cancer with isoforms of a tumor suppressor.

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

GENOMICS

DNA sequencing to support conservation

A huge number of fungal, plant, and animal species are threatened with extinction, and this biodiversity loss is largely caused by human activity. Conservation efforts to protect endangered species are complex and could be supported by genomic tools. In a Perspective, Paez *et al.* discuss how high-quality reference genomes could be used to aid species preservation and population diversity. Additionally, these efforts might be used in de-extinction measures such as genetic rescue to avoid deleterious changes in dwindling populations and even in resurrecting extinct traits. Although genomics will not solve biodiversity loss on its own, these tools could provide important information and new avenues to explore in multifaceted conservation approaches. —GKA

Science, abm8127, this issue p. 364

NANOMEDICINE

A systematic view of nanoparticles

Nanoparticles are increasingly being tested as vehicles for delivering therapeutics, and some are already in clinical use for cancer chemotherapy. Nanoparticle-based treatments can offer various therapeutic advantages such as decreased toxicity, longer half-life, and improved drug delivery. However, there are a multitude of possible nanoparticle formulations with different physical and biological properties, and it is not readily apparent which ones would be best in a given disease setting. Boehnke *et al*. developed a high-throughput screening method that allowed them to systematically evaluate the interactions of 35 different nanoparticle types with hundreds of cancer cell lines (see the Perspective by Winter). Using this approach, the authors identified some features of the cancer cells and nanoparticles that could predict successful nanoparticle delivery, which should facilitate further therapeutic advances. —YN

Science, abm5551, this issue p. 384; see also add3666, p. 371

CORONAVIRUS

Glycans in the spotlight

Viral infection of a cell requires a complex series of molecular recognition events, often mediated by glycoproteins and cell-surface glycans. Buchanan *et al*. developed a nuclear magnetic resonance analysis method to better study such interactions and applied it to influenza and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) proteins binding sialoside glycans. For SARS-CoV-2 in particular, they found evidence for a sialoside-binding site in the N-terminal domain of the original B-origin lineage spike protein that was lost in subsequent variants. These results were corroborated by cryo–electron microscopy structures of the glycan-bound spike protein and genetic variation analysis from patients early in the pandemic, which uncovered host factors involved in glycosylation that potentially contributed to variation in disease severity. —MAF

Science, abm3125, this issue p. 385

PLANT SCIENCE

Genetic improvement drives rice yield

Improvements in agricultural productivity could lessen the impact of agriculture on the environment and perhaps supply more food from less land. Working in rice, Wei *et al*. identified a transcription factor that, when overexpressed, has a variety of useful effects (see the Perspective by Kelly). The gene's expression is induced by both light and low-nitrogen status, and it regulates photosynthetic capacity, nitrogen utilization, and flowering time. In field trials, plants overexpressing this gene delivered greater yield, shortened growth duration, and improved nitrogen use efficiency. —PJH

Science, abi8455, this issue p. 386; see also add3882, p. 370

CORONAVIRUS

A shifting landscape for spike

As proteins evolve, the mutational landscape changes because a mutation at one residue position may affect the functional consequences of a mutation at a second position. To explore how this plays out in the evolution of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein, Starr *et al*. measured the effects of all single-amino-acid mutations on binding of the spike protein to the cellular receptor ACE2 in the context of different SARS-CoV-2 variants (Wuhan-Hu-1, Alpha, Beta, Delta, and Eta). They show how evolution of the protein is shaping the possibility for future mutations, for example, allowing mutations that escape antibodies while maintaining binding to ACE2. —VV

Science, abo7896, this issue p. 420

BIOGEOGRAPHY

Tropical mountain biodiversity

Mountains in the tropics are highly biodiverse, often with entirely different sets of species occurring at high and low elevations. This high turnover occurs because species occupy narrower elevational ranges in the tropics than on temperate mountains. Freeman *et al*. used citizen science data from eBird, a global citizen science project, to test two hypotheses for narrower ranges in the tropics. The prevailing hypothesis is that tropical species experience less seasonal variation in temperature and thus have narrower climatic niches, but an alternative view is that ranges may be limited by competition with other species. Freeman *et al*. found that species richness is a better predictor of smaller elevational ranges than climate seasonality, suggesting a larger role of competition in shaping tropical mountain biodiversity than has previously been recognized. —BEL

Science, abl7242, this issue p. 416

CATALYSIS

Channeling water away

Heterogeneous catalytic reactions that produce water as a by-product can be inhibited by its presence on the surface. Fang *et al*. found that for the production of light olefins from syngas (a 2:1 mixture of hydrogen and carbon monoxide) with a cobalt manganese carbide catalyst at 250°C, the addition of the hydrophobic polymer polydivinylbenzene as part of a physical mixture almost doubled the conversion of carbon monoxide (see the Perspective by Ding and Xu). Theoretical models suggest that the polymer formed channels that accelerated water diffusion away from the catalyst. —PDS

Science, abo0356, this issue p. 406; see also adc9414, p. 369

CORONAVIRUS

Delta and Omicron go toe to toe

There is growing evidence that the Omicron variant of concern (VOC) is more transmissible and infectious than previous iterations of severe acute respiratory syndrome coronavirus 2. Yuan *et al*. compared Syrian hamsters exposed to either Omicron or Delta VOCs. Animals infected with Omicron showed lower respiratory tract viral burdens and reduced clinical severity. Nevertheless, Omicron was at least as transmissible

as Delta, if not more so. When animals were challenged with a mixture of both variants, Delta outcompeted Omicron in naïve hamsters. This competitive advantage disappeared, however, in vaccinated animals. Moreover, vaccinated hamsters were better than unvaccinated animals at transmitting Omicron to co-housed companions. This work helps to clarify how Omicron might have gone on to become the predominant strain in populations with high rates of previous infection and vaccination. —STS

Science, abn8939, this issue p. 428

IMMUNOLOGY

Freeing T cells from a sticky situation

T cells must adhere to and release from other cells to migrate into tissues and stimulate antibody production. The integrin lymphocyte function–associated antigen 1 (LFA-1), which promotes T cell adhesion, is activated downstream of phosphoinositide 3-kinase (PI3K), which generates the lipid second messenger phosphatidylinositol 3,4,5-trisphosphate (PIP3). Johansen *et al.* identified the PIP3-binding, GTPase-activating protein RASA3 as an inhibitor of LFA-1. Because of an increase in adhesion, T cells lacking RASA3 were impaired in entering or exiting lymph nodes, and mice with RASA3-deficient T cells had defective responses to immunization. —WW

Sci. Signal. **15**, eabl9169 (2022).

NEUROSCIENCE

Improving motor skill acquisition in older adults

Activities of daily life typically require executing actions in sequential order, often both accurately and at high speed. In the brain, serial actions are automated into smaller groups of co-occurring actions called motor chunks. Using a sequence-tapping paradigm, Maceira-Elvira *et al.* identified early consolidation of spatial properties of the sequence order and motor chunks to maximize the performance of this task. This process appeared to decline in older adults, but it could be partially restored by using non-invasive brain stimulation during training. With this treatment, motor skill acquisition followed a similar pattern as in young adults, in whom speeding up led to increased motor chunk formation, which in turn allowed for the increasing speed. —CK

Sci. Adv. 10.1126/sciadv.abo3505 (2022).

RESEARCH ARTICLE SUMMARY

NANOMEDICINE

Massively parallel pooled screening reveals genomic determinants of nanoparticle delivery

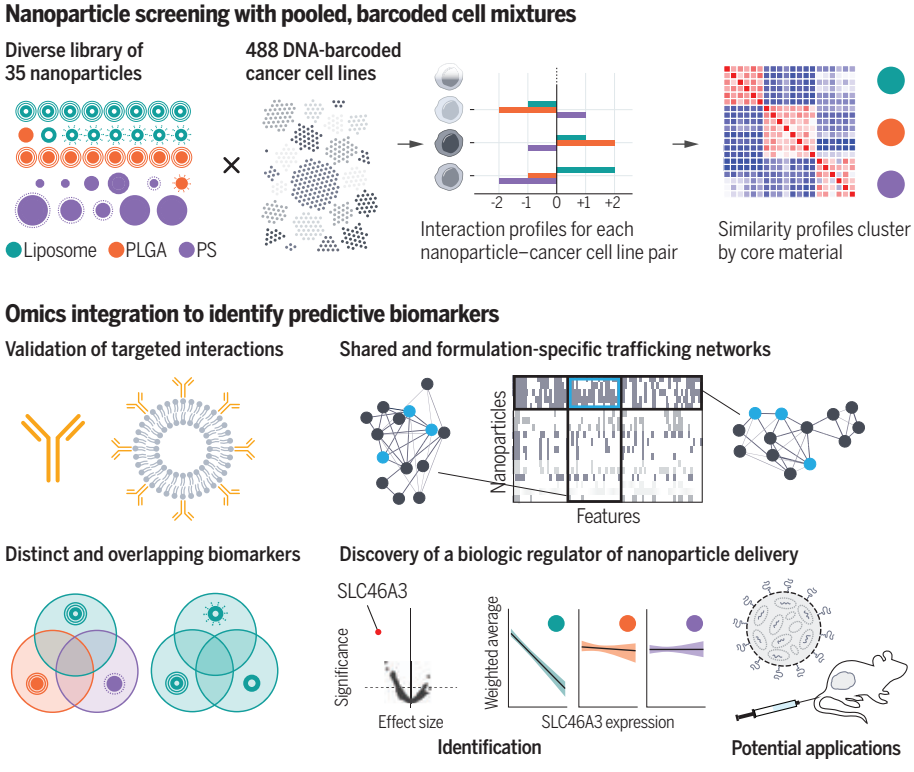
Natalie Boehnke*†, Joelle P. Straehla*†, Hannah C. Safford, Mustafa Kocak, Matthew G. Rees, Melissa Ronan, Danny Rosenberg, Charles H. Adelmann, Raghu R. Chivukula, Namita Nabar, Adam G. Berger, Nicholas G. Lamson, Jaime H. Cheah, Hojun Li, Jennifer A. Roth, Angela N. Koehler, Paula T. Hammond*

INTRODUCTION: Nanoparticle-mediated delivery of therapeutic agents has the potential to considerably affect cancer treatments, particularly in the context of personalized cancer therapies. Nanoparticles span a diverse range of materials and properties. They can be tailored to encapsulate and protect a wide range of therapeutic cargos, including small molecules, biologics, and nucleic acids. One challenge to successful targeted nanoparticle delivery is an incomplete understanding of nano-bio interactions at the target delivery site. In designing this screen, we sought to gain a holistic understanding of both the materials properties and biological features that mediate successful nanoparticle delivery.

RATIONALE: Traditional nanoparticle screens are designed to optimize materials properties in isolated biological contexts. In the era of precision medicine, with the desire to deliver molecularly targeted therapies to specific subcellular compartments, it is also important to probe the structure-function relationship of nanoparticles as they relate to cellular and biological heterogeneity. The combination of pooled screening with multiomic annotation has the potential to accelerate target discovery and uncover previously unrecognized regulators of nanoparticle delivery.

RESULTS: We designed a modular fluorescent nanoparticle library to capture the effects of a wide range of nanoparticle parameters—including core composition, surface chemistry, and size—on cancer cell interactions. In a competitive assay, we screened interactions of each nanoparticle formulation with 488 pooled, barcoded cancer cell lines across 22 lineages and identified cells by strength of nanoparticle association. The screen was designed to probe nanocarrier delivery; thus, no toxic cargo was incorporated. Unsupervised hierarchical clustering of resulting interaction profiles for each nanoparticle–cell line pair identified core composition as a strong determinant of cell association, with the three tested core materials forming distinct clusters. To probe cellular features that govern nanoparticle association, we integrated multiomic data by using correlative analyses. We appropriately identified high epidermal growth factor receptor (EGFR) gene expression and protein abundance as biomarkers that are predictive of cellular affinity for anti-EGFR formulations. More generally, we observed that nanoparticle core material as well as surface modification influence the number and significance of biomarkers that are predictive of uptake. Many biomarkers were associated with established uptake, transport, and adhesion gene sets. Using machine learning algorithms, we identified predictive biomarkers that cluster to form interrelated protein-protein interaction networks, identifying cellular features that mediate nanoparticle trafficking. We also identified formulation-specific biomarkers. We validated expression of *SLC46A3*, a lysosomal transporter, as a negative regulator and predictive biomarker for lipid-based nanoparticle uptake and downstream functional applications. These applications include translation of our findings to in vivo models and demonstration of *SLC46A3* expression modulating both lipid nanoparticle uptake and transfection efficacy of nucleic acid cargo.

CONCLUSION: This work represents a high-throughput interrogation of nanoparticle–cancer cell interactions through the lens of multiomics. Our analysis provides a framework that will empower studies of nano-bio interactions and advance the rational design of nanocarriers. ■



nanoPRISM screen integrates drug delivery and omics. Using a curated nanoparticle library, we screened nanoparticle-cell interaction profiles of hundreds of cancer cells simultaneously. By incorporating omics annotation, we identified biological features, or biomarkers, that mediate nanoparticle delivery to cells. We generated trafficking networks and discovered a biologic regulator of lipid-based nanoparticle delivery. PLGA, polylactide-*co*-glycolide; PS, polystyrene.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: nboehnke@mit.edu, nboehnke@umn.edu (N.B.); jstraehl@mit.edu (J.P.S.); hammond@mit.edu (P.T.H.)
†These authors contributed equally to this work.
Cite this article as N. Boehnke *et al.*, *Science* **377**, eabm5551 (2022). DOI: 10.1126/science.abm5551

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abm5551

RESEARCH ARTICLE

NANOMEDICINE

Massively parallel pooled screening reveals genomic determinants of nanoparticle delivery

Natalie Boehnke[1,2]*†‡, Joelle P. Straehla[1,2,3,4]*†, Hannah C. Safford[1], Mustafa Kocak[2], Matthew G. Rees[2], Melissa Ronan[2], Danny Rosenberg[2], Charles H. Adelmann[5,6,7], Raghu R. Chivukula[6,8], Namita Nabar[1,9], Adam G. Berger[1,10,11], Nicholas G. Lamson[1], Jaime H. Cheah[1], Hojun Li[1,3,4], Jennifer A. Roth[2], Angela N. Koehler[1,2,12], Paula T. Hammond[1,9,10]*

To accelerate the translation of cancer nanomedicine, we used an integrated genomic approach to improve our understanding of the cellular processes that govern nanoparticle trafficking. We developed a massively parallel screen that leverages barcoded, pooled cancer cell lines annotated with multiomic data to investigate cell association patterns across a nanoparticle library spanning a range of formulations with clinical potential. We identified both materials properties and cell-intrinsic features that mediate nanoparticle-cell association. Using machine learning algorithms, we constructed genomic nanoparticle trafficking networks and identified nanoparticle-specific biomarkers. We validated one such biomarker: gene expression of *SLC46A3*, which inversely predicts lipid-based nanoparticle uptake in vitro and in vivo. Our work establishes the power of integrated screens for nanoparticle delivery and enables the identification and utilization of biomarkers to rationally design nanoformulations.

Nanoparticle (NP)–based therapeutics have enormous potential for personalized cancer therapy because they can encapsulate a range of therapeutic cargos, including small molecules, biologics, and more recently, nucleic acids. Therapy-loaded NPs can be designed to prevent undesired degradation of the cargo, increase circulation time, and direct drugs specifically to target tumors (*1*–*3*). There have been notable successes in clinical translation of nanomedicines, including liposomal formulations of doxorubicin (Doxil) and irinotecan (Onivyde) (*4*). These formulations extend the half-life of the active agent and have the potential to lower toxicity but do not efficiently accumulate in tumors (*5*, *6*).

Delivery challenges attributed to circulation, immune detection, and clearance, as well as extravasation and diffusion through tissue, all influence NP accumulation at target disease sites. Efforts to improve NP accumulation in tumors through active targeting motifs have been met with limited success, both in the laboratory and the clinic (*1*, *7*). Fewer efforts have focused on gaining a fundamental understanding of the biological features that mediate successful NP-cell interaction and uptake. Although progress has been made in understanding how specific physical and chemical NP properties affect trafficking and uptake, comprehensive evaluation of multiple NP parameters in combination has thus far been elusive. Additionally, the biologic diversity of cancer targets makes it prohibitively challenging to gain a holistic understanding of which NP properties dictate successful trafficking and drug delivery (*8*, *9*). Once NP parameters are considered in combination, the number of formulations to test increases exponentially, particularly because comparisons across several systems need to be drawn. A further barrier is the need to adapt the NP formulation of each encapsulated therapy for a given drug or target because each formulation has its own distinct biological fate (*9*). As therapies continue to increase in molecular complexity, new nanocarrier formulations capable of delivering such entities will need to be developed and examined for their specific trafficking properties.

We and others have designed panels of NPs to elucidate the structure-function relationships to cellular targeting and uptake (*10*–*13*). However, there is a need to equally consider the influence of biological heterogeneity on interactions at the NP-cell interface—for example, by probing cells across cancer cell lineages with a range of genetic drivers and cell states. In the era of precision medicine, with the desire to deliver molecularly targeted and gene-based therapies to specific subcellular compartments within cancer cells, it is imperative to holistically probe the structure-function relationship of NPs as they relate to cellular interactions.

Inspired by recent advancements in cancer genomics (*14*), we postulated that applying similar techniques to the study of cancer nanomedicine would uncover both the cell- and NP-specific features that mediate efficient targeting and delivery. The combination of pooled screening with multiomic annotation has accelerated target discovery and uncovered previously unrecognized mechanisms of action in small-molecule screens. Specifically, in the profiling relative inhibition simultaneously in mixtures (PRISM) method, DNA-barcoded mixtures of cells have recently been used for multiplexed viability screening. In cell line pools grouped by doubling time, 500 barcoded cell lines have been screened against tens of thousands of compounds to identify genotype-specific cancer vulnerabilities (*15*, *16*).

To comprehensively capture pan-cancer complexities and enable the statistical power to link NP association with cell-intrinsic characteristics, we developed a competitive phenotypic screen to assess associations of a curated NP library with hundreds of cancer cell lines simultaneously. NP-cell association was correlated with genomic features to identify candidate biomarkers. Coupling our biomarker findings with *k*-means clustering, we constructed genomic interaction networks associated with NP engagement, which enabled the identification of genes associated with the binding, recognition, and subcellular trafficking of distinct NP formulations. Moreover, through the use of univariate analyses and random forest algorithms, we identified that the gene *SLC46A3* holds value as a predictive, NP-specific biomarker. We further validated SLC46A3 as a negative regulator of liposomal NP uptake in vitro and in vivo. The strategy outlined here identifies cellular features underlying NP engagement in cancer nanomedicine.

Results

nanoPRISM: Screening NP association with pooled cell lines

To screen hundreds of cancer cell lines simultaneously for NP–cancer cell line association patterns, we cultured pooled PRISM cells and incubated them with fluorescent NPs. We then

[1]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. [2]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [3]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA. [4]Division of Pediatric Hematology/Oncology, Boston Children's Hospital, Boston, MA 02115, USA. [5]Cutaneous Biology Research Center, Massachusetts General Hospital Department of Dermatology, Harvard Medical School, Boston, MA 02114, USA. [6]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA. [7]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. [8]Division of Pulmonary and Critical Care Medicine, Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA. [9]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. [10]Institute for Soldier Nanotechnologies, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [11]Harvard-MIT Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [12]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02142, USA.
***Corresponding author. Email: nboehnke@mit.edu, nboehnke@umn.edu (N.B.); jstraehl@mit.edu (J.P.S.); hammond@mit.edu (P.T.H.)** †These authors contributed equally to this work. ‡Present address: Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455, USA.

implemented a fluorescence-activated cell sorting (FACS) adaptive gating strategy to sort cell populations into four bins (quartiles, A to D) on the basis of fluorescence signal as a proxy for the extent of NP-cell association (Fig. 1A). Experimental parameters were optimized to ensure sufficient cell number and barcode representationafter cell sorting (fig. S1), and NPs were incubated for 4 and 24 hours.

For this screen, we designed a modular NP library to capture the effects of NP core composition, surface chemistry, and size on cell interactions. This panel of 35 NPs encompassed both clinical and experimental formulations. Specifically, anionic liposomes were formulated and electrostatically coated with cationic poly-L-arginine (PLR) followed by a series of polyanions (*17*–*21*). The polyanions were selected for their synthetic [polyacrylic acid (PAA)], semisynthetic [poly-L-aspartate (PLD) and poly-L-glutamate (PLE)], or natural [hyaluronate (HA), dextran sulfate (DXS), fucoidan (FUC), alginate (ALG), and chondroitin sulfate (CS)] origin as well as the inclusion of both carboxylate and sulfate ions (*22*–*24*). These same electrostatic coatings were used to modify polymeric NP cores [polylactide-*co*-glycolide (PLGA)] to test the effects of core composition on NP-cell interactions. We optimized formulations to obtain a diameter of ~100 nm for the liposome and PLGA formulations because the similar sizes would enable cross-core comparisons. We also included commercially manufactured fluorescent carboxylate- and sulfate-modified polystyrene (PS) NPs in a range of diameters from 20 to 200 nm, which enabled the study of particle size and surface chemistry. Because of the clinical importance of polyethylene glycol (PEG)–containing formulations (*25*), PEGylated versions of liposome, PLGA, and PS particles were prepared, including the drug-free versions of two commercial formulations, liposomal doxorubicin (Doxil) and liposomal irinotecan (Onyvide). The latter two formulations are denoted as LIPO-5%

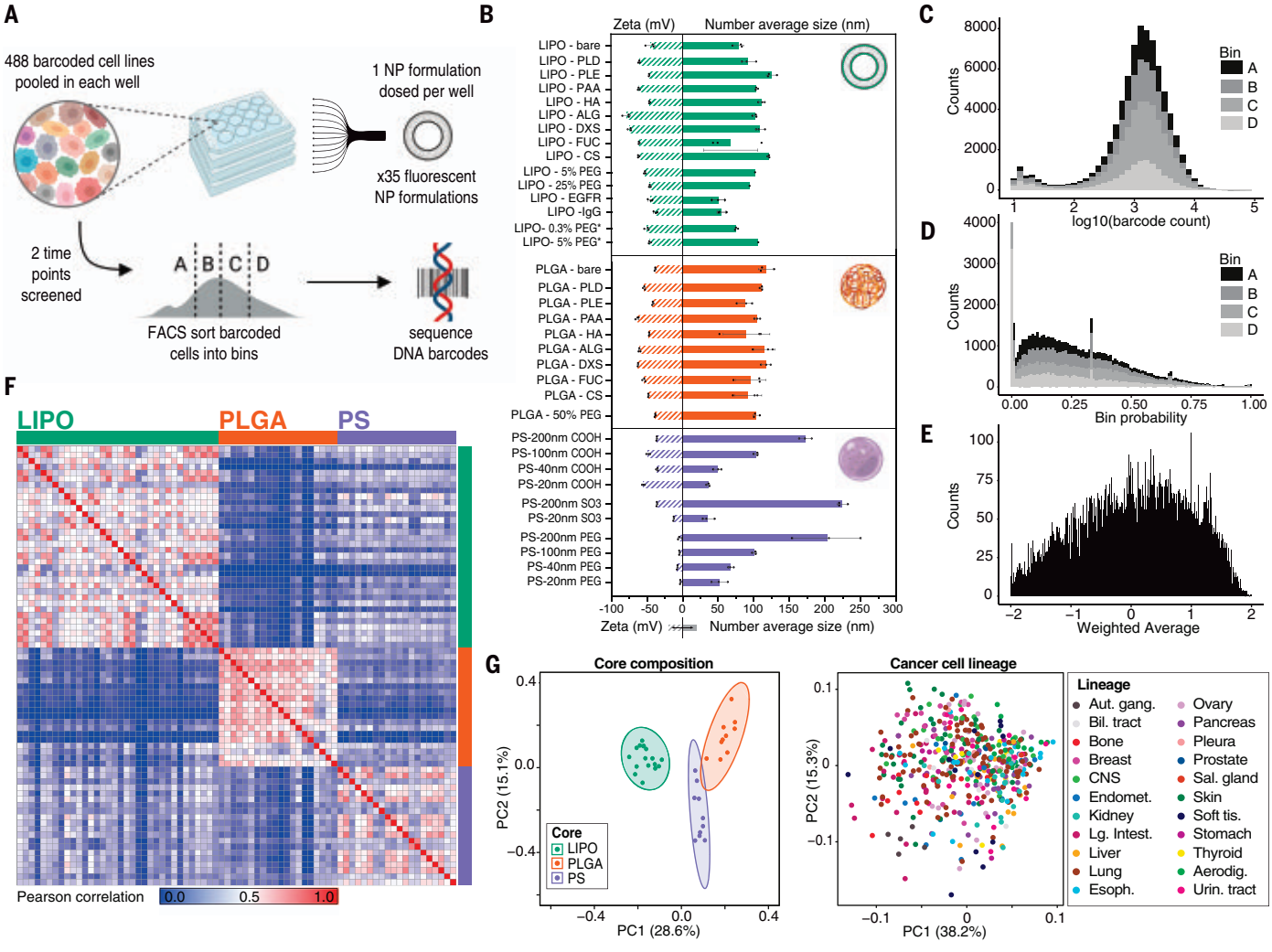


Fig. 1. Assessing NP-cell interactions across hundreds of cancer cell lines simultaneously. (**A**) Schematic of the nanoPRISM assay. Fluorescently labeled NPs are incubated with pooled cancer cells before FACS through NP association and sequencing of DNA barcodes for downstream analyses. (**B**) Characterization of the diameter and ζ potential of the NP library by means of dynamic light scattering. Data are represented as the mean and standard deviation of three technical repeats. Formulations marked with an asterisk indicate drug-free analogs of clinical liposomal formulations as described in the text. (**C**) Raw data from the screen were obtained in the form of barcode counts, with similar numerical distribution of barcodes in each bin, represented as a stacked histogram. (**D**) Accounting for baseline differences in barcode representation yields the probability (*P*) that each cell line will be found in a particular bin. (**E**) Probabilities are collapsed into a single WA for each NP–cell line pair. (**F**) A similarity matrix collapsing WA values for 488 cell lines reveals clusters of NP formulations with the same core formulation. (**G**) PCA of NP–cell line WA values at 24 hours confirms distinct clustering of NP formulations based on core composition (left), but cell lines do not form clusters (right).

PEG* and LIPO-0.3% PEG*, respectively. All of the NPs examined exhibited negative or neutral net charge because the focus of this work is on systemic NP delivery systems. Positively charged NPs have been shown to undergo nonspecific charge interactions with cells and proteins, leading to toxicity and premature clearance in vivo (*26*). Dynamic light scattering (DLS) was used to characterize the diameter, ζ potential, and polydispersity index (Fig. 1B and tables S1 to S3) of this NP library.

To ensure that our methods led to robust and meaningful data, we selected an antibody to epidermal growth factor receptor (EGFR) as an active targeting control. We hypothesized that the design of our screen would allow us to identify features relevant to EGFR expression with a high level of confidence. A nonlethal EGFR antibody or immunoglobulin G (IgG) isotype control was covalently incorporated onto a liposome by means of a PEG tether (*27*). We elected to focus on EGFR because of the wide range of native EGFR expression of the 488 cell lines included in our screen as well as prior evaluation of EGFR-targeting compounds with the PRISM assay (fig. S2) (*15*).

After incubating the cells with the NP library, we used FACS to bin cells into quartiles according to fluorescence intensity (fig. S3). Cells were then lysed, and the DNA barcodes were amplified, sequenced, and deconvoluted according to previously detailed protocols (*15*, *28*). After quality control analysis of technical ($n = 2$) and biological ($n = 3$) replicates, all 488 cell lines met quality control measures and were carried forward for downstream analyses (fig. S4). This dynamic gating strategy was used to enable comparison of cell line representation per bin (quartile) independent of fluorophore identity or amount incorporated into each tested formulation.

A probabilistic model was developed and applied to the data to infer the relative distribution of each cell line into the predetermined bins (A to D) for each NP formulation. The probability of a cell from a given cell line falling into a given bin is used to represent those distributions: $P_A + P_B + P_C + P_D = 1$ (Fig. 1, C and D). The technical details and the model's implementation are presented in the supplementary materials (*29*). Given the concordance of the inferred probabilities among the biologic replicates (fig. S5), we collapsed the replicates through their arithmetic average. Probabilities were then summarized by using a weighting factor alpha (α) to calculate a weighted average (WA) for each NP–cell line pair: $WA = -\alpha P_A - P_B + P_C + \alpha P_D$, in which a higher WA implies higher NP-cell association and vice versa (Fig. 1E). We trialed a range of weighting factors (α = 2, 10, 20, and 100) and found that downstream results were unchanged with the higher α values (fig. S6), and therefore, α = 2 was used for subsequent analyses.

Cancer cells distinguish NPs on the basis of core composition

Pearson-based unsupervised hierarchical clustering of pairwise WAs identified NP core material as a strong determinant of cell association, with the three core materials tested (liposomal, PLGA, and PS) forming distinct clusters (Fig. 1F and fig. S7A). This result was unexpected because we hypothesized surface chemistry to be a larger predictor of NP-cell interactions. Principal components analysis (PCA) similarly identified core-specific trends at both the 4- and 24-hour time points (Fig. 1G and fig. S7, B and C) Further analysis within each core material did reveal surface chemistry–dependent trends, although they were more subtle than core-based clustering (fig. S8).

By contrast, no clusters were apparent when PCA was performed according to cell line, indicating that cancer cells of the same lineage did not have similar NP-association trends (Fig. 1H and fig. S7, B and C). Heterogeneity in NP-cell association in proliferating cells has been attributed to various aspects of cell growth and metabolism (*30*–*33*). To ensure that differential cell proliferation did not confound our results, we performed a parallel growth experiment with the same pooled cells and found no correlation between estimated doubling time and WA (fig. S9).

Cell-intrinsic features mediate NP trafficking

We applied data from the Cancer Cell Line Encyclopedia (CCLE) (*34*, *35*) to identify genomic features that act as predictive biomarkers for NP-cell association. To do this, we used both univariate analyses and a random forest algorithm to correlate the baseline molecular features of each cell line (cell lineage; gene copy number; mRNA, microRNA, protein, or metabolite abundance; and function-damaging, hotspot, or missense mutations) with NP association (fig. S10, A and B).

EGFR-targeting compounds identified relevant biomarkers with high confidence

Using univariate analysis for all CCLE features, we identified EGFR gene expression and protein abundance as the two most significantly correlated hits ($q = 4 \times 10^{-100}$ and $q = 4 \times 10^{-76}$, respectively) with antibody to EGFR, but much less significantly ($q = 6 \times 10^{-9}$ and $q = 4 \times 10^{-10}$, respectively) associated with the isotype control (Fig. 2A, top). We also confirmed that fluorophore identity does not affect biomarker identification, demonstrating that both AlexaFluor 488– and Cy5-conjugated antibodies to EGFR perform similarly (fig. S10, C to E).

In EGFR-conjugated liposomes, the same hits were also identified more significantly ($q = 6 \times 10^{-21}$ and $q = 2 \times 10^{-18}$, respectively) than the those of IgG control ($q = 3 \times 10^{-9}$ and $q = 3 \times 10^{-6}$, respectively) (Fig. 2A, bottom).

The statistical significance of EGFR biomarkers was lower for the antibody-conjugated liposome than the free antibody, which may be due to changes in protein concentration across samples or steric blockage introduced by covalently linking an antibody to a NP surface that may interfere with binding to its target (*36*). Thus, we demonstrated the ability to quantitatively compare expected biomarker targets of both free antibodies and antibody-conjugated NPs by using our platform.

Biomarker number and identity are influenced by NP properties

We applied univariate analysis to correlate association and CCLE features for each NP formulation both quantitatively and qualitatively by using curated gene sets. First, we thresholded q values at less than 1×10^{-10} to compare the absolute number of candidate biomarkers at varying degrees of significance (Fig. 2B). Selection of this cutoff was guided by the IgG-conjugated antibody analysis, which returned few hits above this threshold. For liposomal NPs, we observed that the number of significant biomarkers was higher at 4 hours than 24 hours. We believe that this may be indicative of active uptake processes, established to take place within the first few hours of NP-cell interactions, whereas at 24 hours, we may be capturing features associated with less specific interactions (*37*, *38*). We next investigated biomarkers associated with established uptake, transport, and adhesion gene sets (Fig. 2C) (*39*–*41*). To examine the distribution of biomarker significance across curated gene sets and NP formulations, each gene was visualized by using the –log(q value) for gene expression. As expected, we identified highly significant biomarkers from gene sets important in drug import and export such as solute carrier (SLC) transporter family and adenosine 5′-triphosphate (ATP)–binding cassette (ABC) family. Our screen provides data on both the significance and the relationship to NP delivery. For example, we found that ABCA1, which plays a role in cholesterol transport, has a positive relationship with liposomal NPs, whereas several members of the multidrug-resistance subfamily (ABCB1/P-GP, ABCC1/MRP, and ABCC4/MRP4) have a negative relationship with PLGA NPs (fig. S11) (*42*). We also identified biomarkers important for cell engagement (focal adhesion and extracellular matrix) as well as intracellular trafficking (vesicular transport, lysosome, and cholesterol transport). This highlights the ability of our screen to identify expected biomarkers and enable comparison between drug delivery modalities.

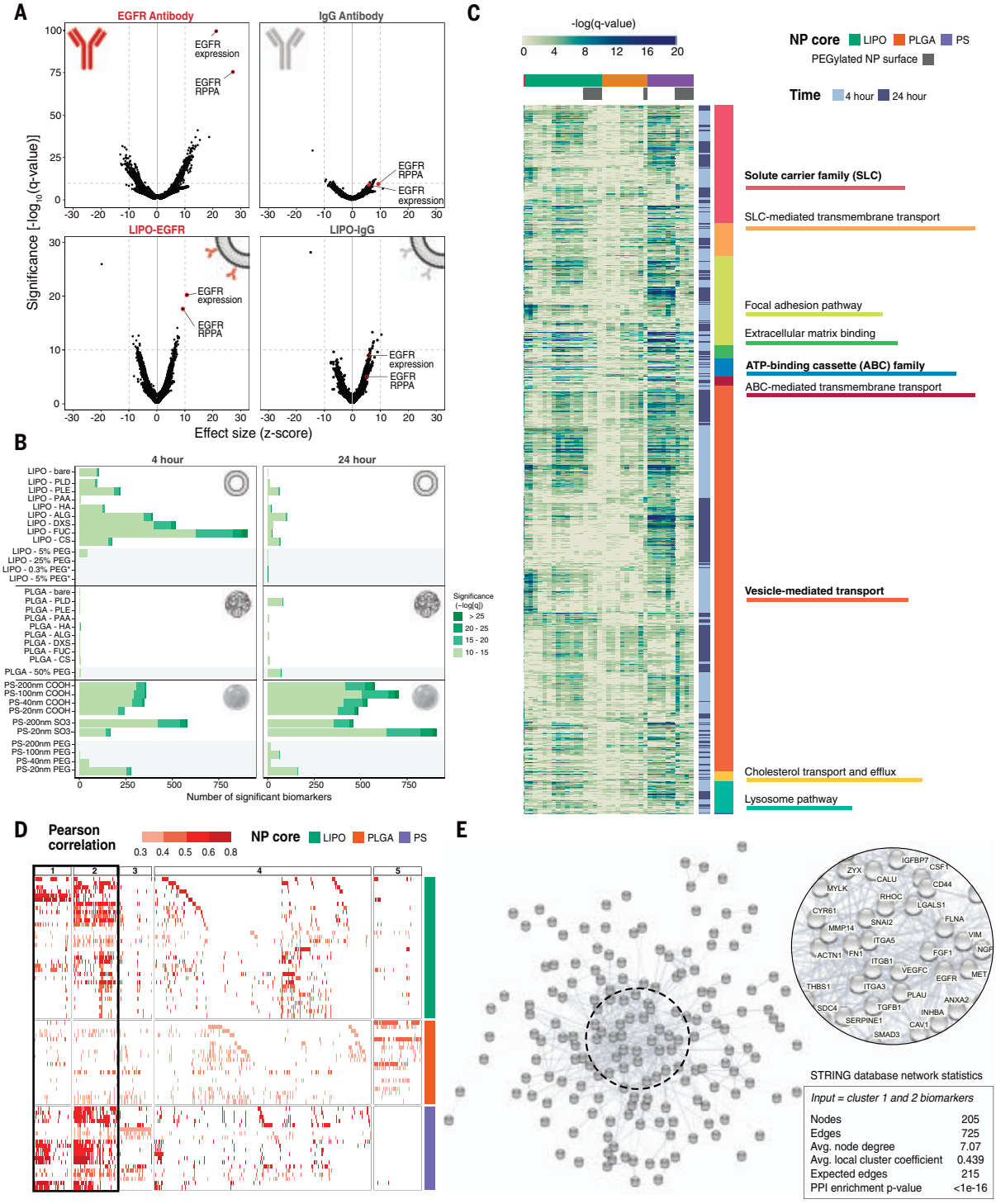


Fig. 2. Correlative genomic analysis identifies expected validation biomarkers as well as hundreds of formulation- and time-dependent biomarkers. (**A**) Cells with high EGFR antibody association are strongly correlated with EGFR gene expression and protein abundance [by means of reverse phase protein array (RPPA)] (top left). These correlations are diminished in the isotype control-treated sample. The same EGFR-related hits, in addition to NP-specific markers, are observed for cells treated with antibody-conjugated liposomes (bottom row). (**B**) Univariate analysis identifies genomic features correlated with NP association. All biomarkers meeting a significance threshold of –log10(*q* value) > 10 are shown as stacked bar graphs separated by NP formulation and time point. PEGylated NP formulations are indicated with a gray background. (**C**) A heatmap showing the significance of biomarkers associated with established transport, uptake, and adhesion gene sets. Gene set headings are bolded, and subsections are listed below respective headings. (**D**) A heatmap showing all gene- and protein-expression features, with positive correlation identified by means of random forest algorithm in columns, and NP formulations in rows. Features are colored on the basis of their Pearson correlation and clustered by using *k*-means clustering, with clusters 1+2 highlighted as features present across multiple NP formulations. (**E**) Visual representation of the STRING network generated by inputting the 205 features from clusters 1+2, with network statistics. Each node indicates a feature, and the edges indicate predicted functional associations. (Inset) Zoom-in with the most interconnected nodes labeled.

We also observed that liposome surface modification influences the number and significance of biomarkers. Specifically, liposomes electrostatically coated with polysaccharides (HA, ALG, DXS, FUC, and CS) had the highest amount of associated biomarkers, which we hypothesize is due to the high degree of interactions between sugars and cell surface proteins as well as the potential for naturally occurring polysaccharides to interact with a wide range of cell surface elements (*23*, *43*, *44*). In line with this hypothesis, the addition of PEG, a well-established antifouling polymer, reduces the number and significance of associated biomarkers almost to zero. In light of the highly specific hits generated from EGFR-conjugated liposomes (formulated by using 25% PEG liposomes), this abrupt decrease in significant biomarkers further indicates the ability of our platform to identify specific NP binding and recognition elements. In contrast to the liposomal formulations, PLGA formulations, regardless of surface modification, resulted in few biomarkers at either time point. Last, a high number of significant biomarkers was associated with both carboxylated and sulfated PS NPs included in our screen, although there was no time dependence, in contrast to the liposomal formulation. Although this result was unexpected, because the PS formulations are made of synthetic polystyrene polymers, meaningful biological interactions with anionic polystyrenes both in polymer and particle form have been reported. Specifically, it was described that NPs bearing anionic polystyrene motifs have the appropriate mix of hydrophobicity and anionic charge character to interact favorably with trafficking proteins, including the caveolins (*45*).

NP biomarkers are connected and create trafficking networks

We then used an unbiased approach to identify predictive biomarkers by using a random-forest algorithm, annotated by feature set: gene expression, gene copy number, and protein abundance. Data from the 4-hour time point were chosen for this analysis on the basis of the EGFR-related hits for liposomes, which were more significant at 4 hours than at 24 hours. Because we were interested in applying this approach to identify cellular features positively correlated with uptake (for example, increased expression of trafficking proteins), hits negatively correlated with NP association were removed from this analysis. Next, we used *k*-means clustering to visualize biomarkers according to their relative importance and presence across formulations (Fig. 2D). Clusters 1 and 2 contained 205 hits shared across NP formulations and were especially enriched for liposomal and PS NPs. These genes and proteins were input into the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (*46*–*48*) to generate a protein-protein interaction (PPI) network that was found to be highly interconnected (PPI enrichment $P < 1 \times 10^{-16}$) (Fig. 2E). The network is enriched in proteins found in the plasma membrane, extracellular region, and extracellular matrix [false discovery rate (FDR) = 8×10^{-12}, 3×10^{-9}, and 3×10^{-8}, respectively] on the basis of enrichment analysis with Gene Ontology (GO) localization datasets (fig. S12) (*49*–*51*). The identification of overlapping biomarkers that are localized to the cell surface and have established protein-protein interactions led us to hypothesize that these proteins are important in early NP trafficking. Enrichment analyses by using GO molecular functions datasets showed enrichment in numerous binding processes (data S1 and fig. S12), giving further credence to this theory.

SLC46A3 is a negative regulator of liposomal NP uptake

Evaluating univariate results across NP formulations, we identified one biomarker with a strong, inverse relationship with liposomal NP association: expression of solute carrier family 46 member 3 (*SLC46A3*). A member of the solute carrier (SLC) transporter family, SLC46A3 is a relatively unstudied transporter that has been localized to the lysosome (*52*, *53*). SLC46A3 was recently identified as a modulator of cytosolic copper homeostasis in hepatocytes, connecting hepatic copper concentrations with lipid catabolism and mitochondrial function (*54*). This reported relationship between SLC46A3 and lipid catabolism may help to explain why SLC46A3 was found to have a strong relationship with liposomal NP uptake and not uptake of polymeric NPs. In the context of cancer, SLC46A3 was recently shown to transport noncleavable antibody-drug conjugate (ADC) catabolites from the lysosome to the cytosol, thereby being necessary for therapeutic efficacy (*55*). Further, down-regulation of SLC46A3 was identified as a resistance mechanism for ADC delivery in cancer cells, including in patient samples of multiple myeloma (*55*–*58*). Although the biologic function of SLC46A3 in cancer is not yet clear, given the potential therapeutic implications and the unusual inverse relationship between SLC46A3 expression and NP delivery, we sought to validate the predictive power of *SLC46A3* as a biomarker for liposomal NP association.

SLC46A3 expression was the most significant hit on univariate analysis and also the top ranked random forest feature for each liposomal NP tested at 24 hours, regardless of surface modification ($q < 10^{-20}$) (Fig. 3A and fig. S13). This inverse relationship between *SLC46A3* expression and NP association was found to be specific to liposomal NPs, and not observed with PLGA or PS NPs, and was maintained regardless of cancer cell lineage (Fig. 3B and fig. S13).

We selected nine cancer cell lines from the nanoPRISM pool and four additional cell lines, spanning multiple lineages, with a range of native *SLC46A3* expression levels for screening in a nonpooled fashion (Fig. 3, C and D, and figs. S3, S14, and S15). Analogous to the pooled screen, individual cell lines were profiled by using flow cytometry, and NP-associated fluorescence was quantified after 24 hours incubation; *SLC46A3* expression was concurrently quantified by using quantitative polymerase chain reaction (qPCR) (Fig. 3D and fig. S9). In line with observations from pooled screening, the inverse relationship between liposome association and native *SLC46A3* expression was maintained, suggesting that *SLC46A3* may play a key role in regulating the degree of liposomal NP uptake.

To probe whether *SLC46A3* governs cellular association with NPs, we selected the breast cancer cell line T47D, which exhibited high native *SLC46A3* (Fig. 4A). We deactivated *SLC46A3* with small interfering RNA (siRNA) and evaluated the effect on liposomal NP association. We observed that T47D cells with reduced *SLC46A3* had higher NP-cell association with both tested formulations, which suggested that modulating *SLC46A3* expression alone can regulate NP-cell association (Fig. 4B).

To further functionally evaluate the relationship of *SLC46A3* expression and NP-cell association, we selected two cancer cell lines from the pooled screen (Fig. 4A): the T47D cell line and the melanoma cell line LOXIMVI. We developed a toolkit using these two cell lines by permanently deactivating *SLC46A3* in T47D cells and inducing *SLC46A3* overexpression in LOXIMVIs (fig. S16, A to D).

Because SLC46A3 is a protein associated with lysosomal membranes (*55*, *56*, *59*), we used LysoTracker dye to evaluate the effect of *SLC46A3* modulation on endolysosomal compartments in both T47D and LOXIMVI engineered cell lines (Fig. 4C). We observed an *SLC46A3*-dependent change: cells with lower SLC46A3 expression (T47D-*SLC46A3* deactivation, LOXIMVI-vector control) exhibited more brightly dyed endolysosomal compartments as compared with that of their high-SLC46A3-expression counterparts (T47D-vector control, LOXIMVI-SLC46A3 OE).

Overexpression of SLC46A3 in LOXIMVI cells significantly abrogated interaction with bare liposomes ($P = 0.006$) by using flow cytometry profiling (Fig. 4D). The T47D-*SLC46A3* deactivation cell line demonstrated significantly increased association with bare liposomes compared with that of parental or vector control lines ($P = 0.0017$) (Fig. 4D). We further confirmed that these trends are generalizable across a range of surface functionalized

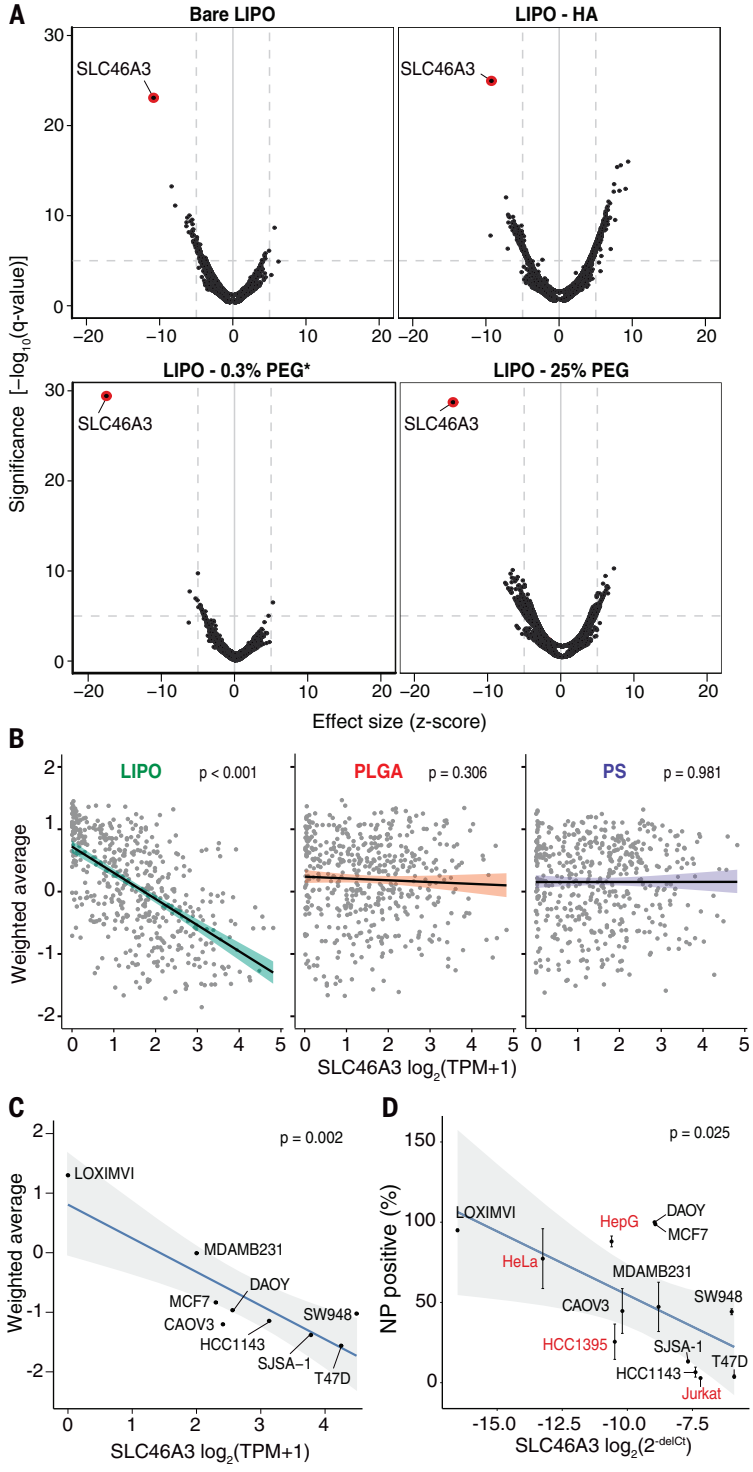


Fig. 3. Native expression of the lysosomal transporter SLC46A3 predicts NP-cell interaction for liposome formulations. (**A**) Univariate analysis identified SLC46A3 expression as strongly inversely correlated with liposome association, regardless of liposomal surface modification. (**B**) Using linear regression to evaluate the biomarker relationship across core formulations revealed that *SLC46A3* expression was inversely correlated with NP association in liposome–cell line pairs ($P < 0.001$) but not PLGA– and PS–cell line pairs ($P > 0.05$); $n = 488$ cell lines for each plot. (**C**) Cell lines in the nanoPRISM pool exhibited a range of native *SLC46A3* expression and a log linear correlation with uptake of bare liposomes. (**D**) This same correlation was also exhibited when assessing liposome-cell associations with flow cytometry in a nonpooled fashion ($P = 0.025$). Cell lines in red were not part of the pooled PRISM screen. Data represented in (D) are shown as the mean and standard deviation of four biological replicates. Error bars are not shown when smaller than data points.

liposomes (Fig. 4E and fig. S16E). Moreover, no significant changes in NP association were observed for PLGA and PS NPs (Fig. 4E and fig. S16, F and G). We also confirmed that the presence of serum proteins in cell culture media does not abrogate this trend (fig. S16H). Taken together, these data indicate that modulation of SLC46A3 alone in cancer cells is sufficient to negatively regulate association and uptake of liposomal NPs.

Because flow cytometry does not provide spatial information with respect to NP-cell interactions, we used imaging cytometry to characterize NP localization in a high-throughput manner (Fig. 5, A to F). We selected four representative formulations: three liposomal NPs to probe the relationship of SLC46A3 expression with liposome trafficking and one PLGA NP formulation with the same outer layer (PLD).

Consistent with trends observed with flow cytometry, we observed an inverse relationship between NP intensity and SLC46A3 expression for liposomal, but not PLGA, NPs (Fig. 5, A and D, and fig. S11). Using brightfield images, we applied a mask to investigate cellular localization of NPs. All tested formulations were internalized, and this did not change with SLC46A3 modulation (Fig. 5, B and E).

We investigated localization of NPs by scoring NP signal according to distribution within each cell (Fig. 5, C and F, and fig. S17). We observed stark differences in median cellular distribution scores of liposomal NPs in relation to SLC46A3 expression in T47D cells. This was not observed for PLGA NPs, mimicking the previously observed core-specific relationship between NP-cell association and SLC46A3 expression. Changes in this score, although less pronounced, were also observed for liposomal NPs in LOXIMVI cells.

To confirm our findings with higher spatial resolution, we used deconvolution microscopy of live cells and incorporated a lysosomal stain to observe changes in intracellular trafficking (Fig. 5, G and H). NPs appeared uniformly distributed within T47D-*SLC46A3*–deactivation cells, colocalizing with endolysosomal vesicles. By contrast, LIPO-PLD NPs were localized to large endolysosomal clusters in T47D-vector control cells. This trend was also observed for LIPO-PLE and LIPO-0.3% PEG* NPs and at the earlier time point of 4 hours (fig. S18). Changes in localization were not observed for the tested PLGA PLD NPs. This again indicates a NP core-dependent relationship with SLC46A3.

In the engineered LOXIMVI cell lines, we also observed colocalization of liposomal NPs with endolysosomal signal (Fig. 5H). However, predictable changes in NP localization were not detected, which is in line with smaller changes in median cellular distribution scores.

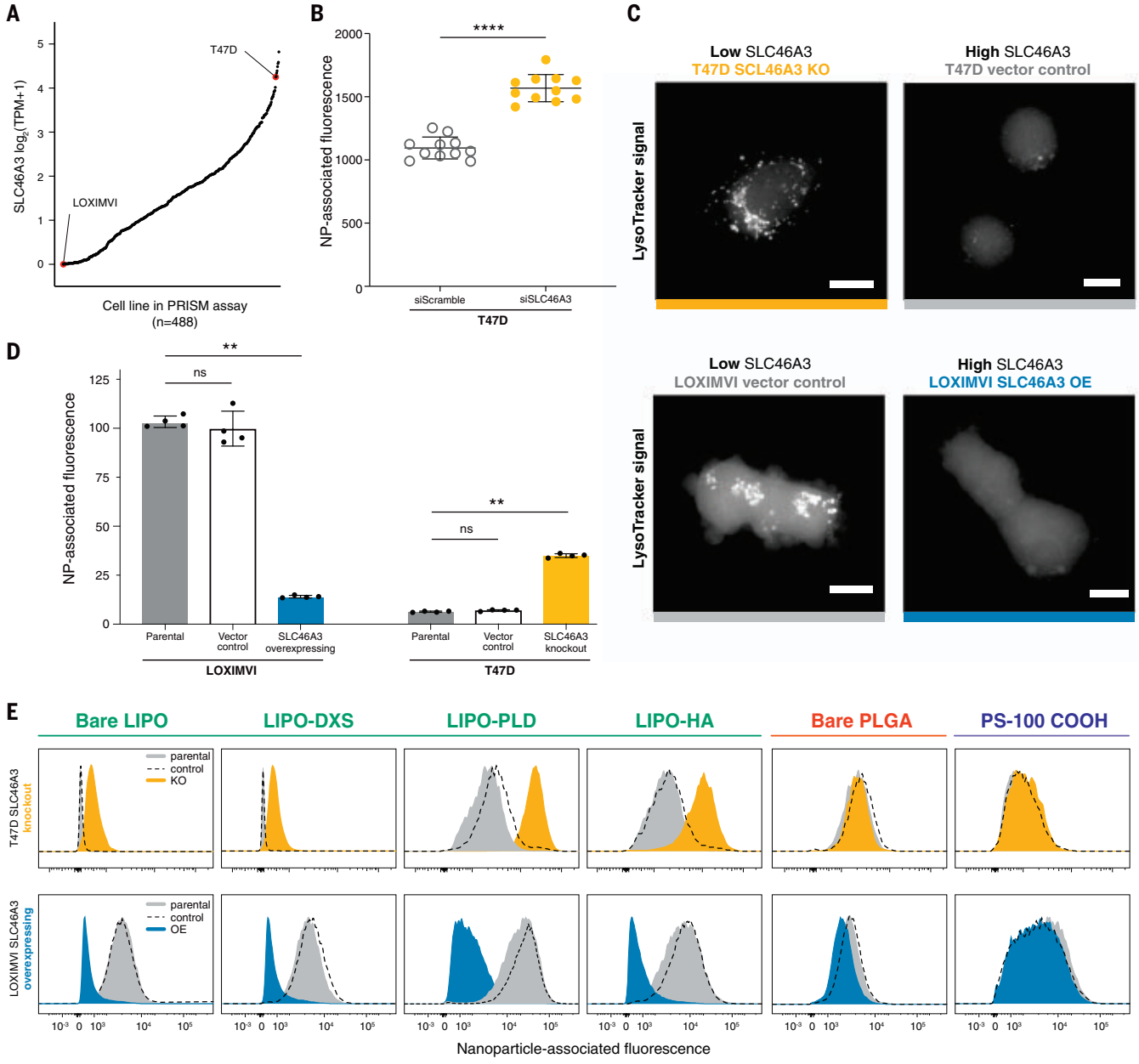


Fig. 4. Modulating *SLC46A3* expression in cancer cell lines is sufficient to negatively regulate interaction with liposome NP formulations. (**A**) T47D and LOXIMVI cells have high and low *SLC46A3* expression, respectively, among the cells in the nanoPRISM cell line pool. (**B**) T47D cells treated with siRNA to deactivate SLC46A3 have higher uptake of Lipo-PLD compared with that of T47D cells treated with a scrambled siRNA control (****$P < 0.0001$, Mann-Whitney test). (**C**) Representative micrographs of Lysotracker signal in engineered cell lines showed endolysosomal compartments. KO, knockout (deactivated). Scale bars, 10 μm. (**D**) Using lentivirus to overexpress *SLC46A3* in LOXIMVI cells and CRISPR/Cas9 to permanently deactivate *SLC46A3* in T47D cells, we showed that modulation results in significantly changed liposome association, as determined with flow cytometry (**$P < 0.001$, Kruskal-Wallis test). NP-associated fluorescence is defined as median fluorescence intensity normalized to untreated cells. Data are represented as the mean and standard deviation of four biological replicates. (**E**) Shifts in NP association were consistently observed across all tested liposomes, independent of surface modification. No shifts were observed with PLGA or PS formulations.

Impact of SLC46A3 expression on endolysosomal maturation is minimal

To further probe the relationship between intracellular liposomal NP trafficking and SLC46A3 expression, we used imaging cytometry to spatially interrogate markers of endolysosomal transport. We elected to study markers of early (EEA1 and RAB5A), late (RAB7), and recycling endosomes (RAB11) as well as lysosomes (LAMP1) in engineered LOXIMVI cells (figs. S19 and S20 and table S4). Although no apparent differences in endolysosomal marker signal strength, size, and shape were observed when comparing LOXIMVI-SLC46A3 OE and LOXIMVI-vector control cells both in the absence and in the presence of liposomal NPs, modest changes in EEA1, RAB7, and LAMP1 texture were noted (fig. S19, A and B).

We then assigned values to the colocalization between each endolysosomal marker and NP signals and observed increasing colocalization from EEA1 to RAB5 to RAB7, which is consistent with liposome trafficking from early to late endosomes (fig. S19, C to F).

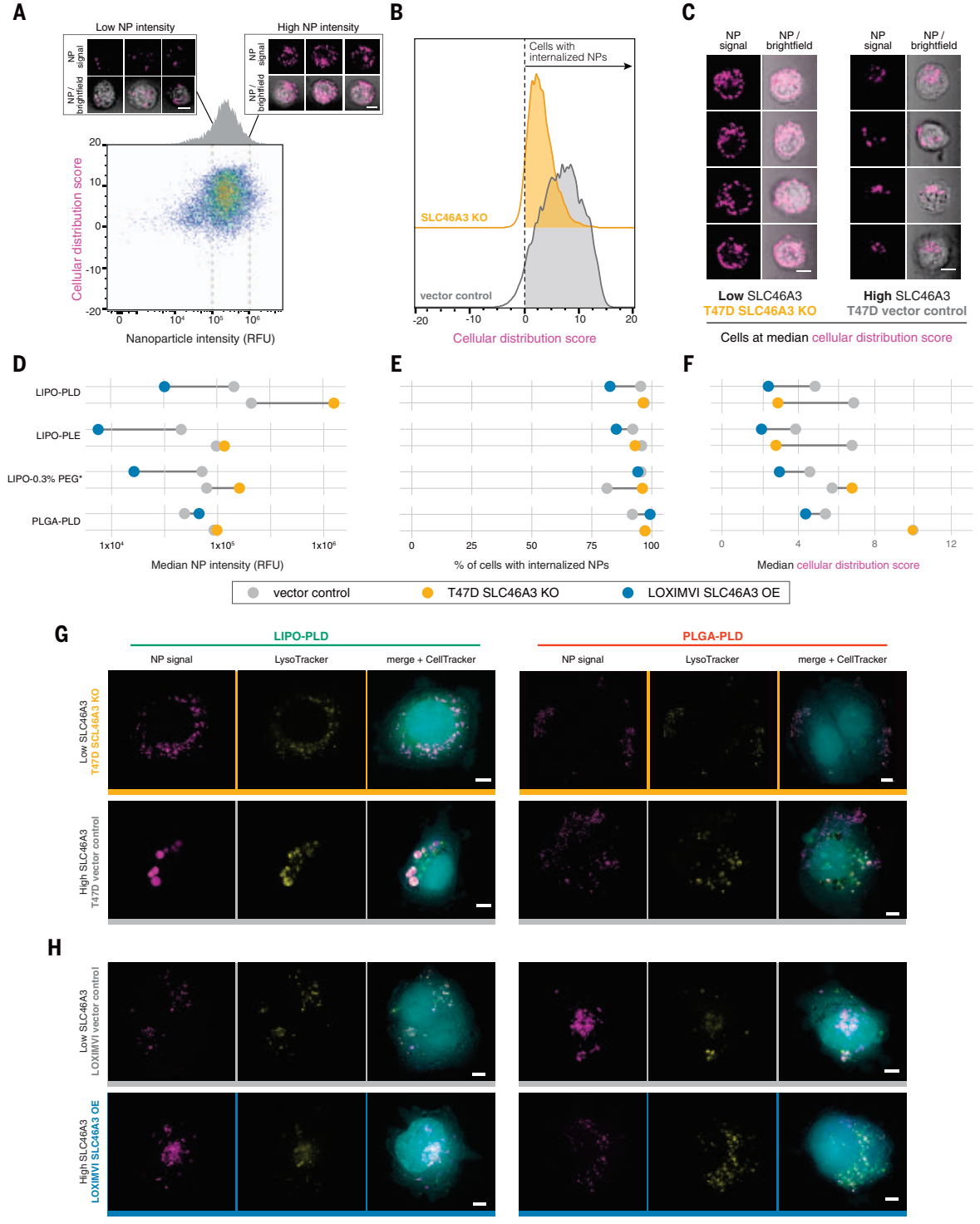


Fig. 5. High-throughput imaging cytometry confirmed NP internalization and revealed SLC46A3-dependent changes to intracellular trafficking. (**A**) Imaging cytometry was used to investigate the intensity (*x* axis) and distribution (*y* axis) of NPs in a high-throughput manner. (Bottom) Bivariate density plot of *n* = 10,000 cells (T47D-vector control) after 24 hours incubation with LIPO-PLD NPs, with representative cell images at low and high NP signal. (**B**) Cellular distribution patterns of NPs were scored so that scores greater than 0 indicate cells with internalized NPs. Representative data from LIPO-PLD NPs in engineered T47D cells are shown. (**C**) Representative cell images at the median cellular distribution score for engineered T47D cells treated with LIPO-PLD NPs. (**D**) Quantification of median intensity of tested NP formulations in engineered T47D and LOXIMVI cell lines demonstrated SLC46A3-dependent changes. (**E**) NPs remained predominantly internalized independent of SLC46A3 expression. (**F**) Shifts in the median cellular distribution scores were observed in response to SLC46A3 modulation. (**G** and **H**) Live cell micrographs of (G) T47D-vector control and T47D-*SLC46A3* deactivation cells and (H) LOXIMVI-vector control and LOXIMVI-SLC46A3 OE cells incubated with LIPO-PLD and PLGA-PLD NPs for 24 hours. NP signal is pseudo-colored magenta, LysoTracker signal is yellow, and CellTracker is cyan. Scale bars, (A) and (C), 7 μm; (G) and (H), 5 μm.

Colocalization between RAB7 and liposomal NPs was higher in LOXIMVI-SLC46A3 OE cells as compared with vector control, and the opposite relationship was observed for LAMP1 colocalization.

Liposome retention and accumulation remains SLC46A3 dependent in vivo

To evaluate the potential clinical utility of SLC46A3 as a negative regulator of liposomal NP delivery, we tested in vivo delivery of a US Food and Drug Administration (FDA)–approved NP analog, the drug-free version of liposomal irinotecan (LIPO-0.3% PEG*), in mice bearing subcutaneous LOXIMVI flank tumors. Fluorescently labeled NPs were administered by means of a one-time intratumoral injection or repeat intravenous administration to evaluate tumor retention and accumulation, respectively (Fig. 6A and fig. S21).

NP signal was quantified at both 4 and 24 hours after intratumoral administration. In line with our hypothesis, as well as with in vitro NP-associated fluorescence data (fig. S21A), we observed an inverse relationship between SLC46A3 expression and LIPO-0.3% PEG* NP retention that became more pronounced over time (P = 0.0115, 4 hours; P = 0.0002, 24 hours) (Fig. 6, B and C, and fig. S21, B to E). Moreover, these findings also align with our initial nanoPRISM findings, in which SLC46A3 expression was a more significant biomarker at 24 hours ($q = 3.49 \times 10^{-30}$) (data S2 and fig. S13A) than at 4 hours ($q = 1.47 \times 10^{-4}$) (data S2 and fig. S13A).

To determine whether *SLC46A3* expression predictably governs accumulation of nontargeted NPs, which bear no specific functional ligands on their surface, after systemic administration, we quantified NP signal after intravenous injections. We observed a significant relationship between SLC46A3 and NP accumulation (P = 0.0019) (Fig. 6D and fig. S21F). This demonstrates that baseline tumor expression of SLC46A3 may influence NP delivery in a physiologic setting.

Together, these data highlight the real-world relevance of the nanoPRISM screening assay in general as well as the utility of SLC46A3 in particular as a potential biomarker.

Solid lipid NP uptake and transfection are dependent on SLC46A3 expression

Given the recent translational success and promising potential of nucleic acid–carrying solid lipid NPs (LNPs) (*60*, *61*), we sought to determine whether the relationship of SLC46A3 expression extends to LNP association as well as transfection efficiency. We generated fluorescently (Cy5) labeled LNPs that contained mRNA encoding green fluorescent protein (GFP) (LNP 1) and incubated these particles with engineered LOXIMVI cell lines (tables S3 and S5).

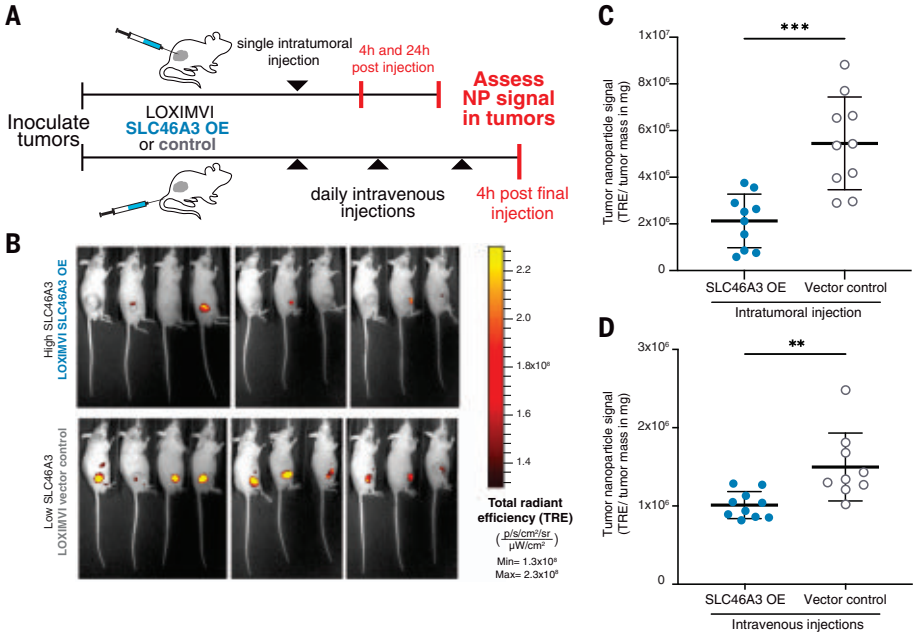


Fig. 6. Retention and accumulation of PEGylated liposomes (LIPO-0.3% PEG*) in LOXIMVI tumors is dependent on SLC46A3 expression. (**A**) Fluorescently labeled LIPO-0.3% PEG* NPs were administered to mice bearing LOXIMVI flank tumors by means of a one-time intratumoral injection or repeat intravenous injections. (**B**) Whole-animal fluorescence images of mice (four males, six females per group) 24 hours after being intratumorally injected with LIPO-0.3% PEG* NPs. (**C**) Quantification of LIPO-0.3% PEG* NP retention 24 hours after intratumoral administration to LOXIMVI flank tumors. (**D**) Quantification of LIPO-0.3% PEG* NP accumulation after repeat intravenous injections. In (C) and (D), NP signal is expressed on the y axis as total radiant efficiency divided by tumor mass; units are provided in the figure. The mean and standard deviation of n = 10 mice are shown with the exception of the LOXIMVI-vector control, repeat intravenous injection group, where n = 9 mice (**P < 0.01, ***P < 0.001, Mann-Whitney test).

LNP association, as quantified by Cy5 signal, was significantly lower for LOXIMVI-SLC46A3 OE cells than LOXIMVI-vector control cells, showing the same relationship (lower SLC46A3 expression correlating with higher association) for LNPs as for liposomal NPs (P = 0.008) (Fig. 7, A and B). A similarly inverse relationship with *SLC46A3* expression was seen for transfection, as quantified by GFP signal of formulation LNP 1 (Fig. 7C). Taken together, these findings suggest that SLC46A3 regulates cytosolic delivery of mRNA cargo by way of LNP uptake. Expanding on this, we generated two additional LNPs, analogous to commercial formulations (table S5) (*62*–*65*). Although we observed lower transfection in LOXIMVI-SLC46A3 OE cells than in LOXIMVI-vector control cells, these differences were not statistically significant (P > 0.05). Nevertheless, the inverse relationship between SLC46A3 expression and cell association in multiple LNP formulations supports the relevance of SLC46A3 as a predictive biomarker for lipid-based NP formulations.

Discussion

This work represents high-throughput interrogation of NP–cancer cell interactions through the lens of multiomics. Harnessing the power of pooled screening and high-throughput sequencing, we developed and validated a platform to identify predictive biomarkers for NP interactions with cancer cells. We used this platform to screen a 35-member NP library against a panel of 488 cancer cell lines. This enabled the comprehensive study and identification of key parameters that mediate NP-cell interactions, highlighting the importance of considering both nanomaterials and cellular features in concert.

Although pooled screening is a powerful tool, there are several important limitations. First, we primarily focused on lipid-based and polymeric NP formulations with translational drug delivery potential. We recognize that there are several additional categories of nanomaterials with wide-ranging properties, such as inorganic systems, that can be useful for both therapeutic and diagnostic applications (*66*, *67*), and we believe that additional biomarkers that mediate the trafficking of inorganic NPs may be identified by using similar screening approaches. Second, the results of in vitro screens are often met with limited success when translated in vivo because NP-mediated delivery is dependent on many

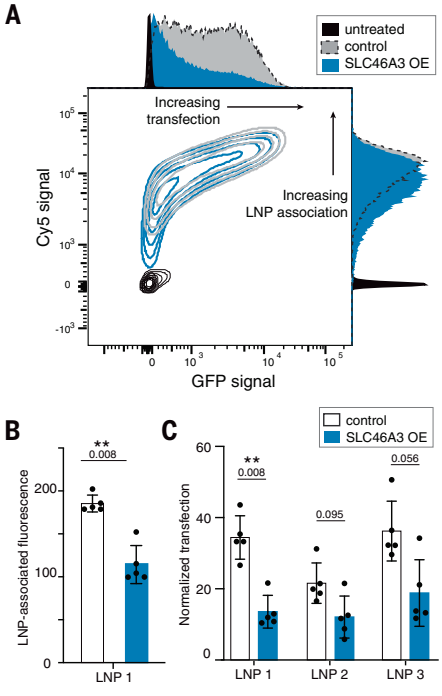


Fig. 7. Solid lipid NP-cell association and transfection are *SLC46A3*-dependent, as determined with flow cytometry. (**A**) Contour plot of Cy5 signal and GFP signal indicating decreased LNP-cell association and transfection efficacy in LOXIMVI cells overexpressing *SLC46A3*. (**B**) Quantification of LNP signal revealed a significant change in LNP-cell association across control and *SLC46A3*-overexpressing LOXIMVI cells (**P = 0.008, Mann-Whitney). LNP-associated fluorescence is defined as median fluorescence intensity normalized to untreated cells. (**C**) LOXIMVI-SLC46A3 OE cells exhibited lower transfection efficiency than that of LOXIMVI-vector control cells after dosing of three different LNP formulations (Mann-Whitney). Normalized transfection is defined as median GFP intensity normalized to untreated cells.

factors beyond the nano-cell interface (*8*). However, the level of molecular characterization and statistical and computational power afforded by annotated biological datasets, such as the CCLE, is currently unrivaled. Therefore, existing in vivo screens cannot yet provide this breadth or statistical power. Keeping translational barriers in mind is key to the successful validation of candidate biomarkers, and for this reason, we used multiple isogenic models and tested a range of lipid-based NPs across in vitro and in vivo conditions. Third, an additional limitation of this screen is related to the availability of genomic datasets for each cell line tested because dataset completeness contributes to the power of detection for both univariate and multivariate analyses. At the time of analysis, 10 feature sets were available for the majority of cell lines in our pool. However, as datasets expand over time, it will be possible to reanalyze our data in the future. Especially for emerging fields such as proteomics and metabolomics, the opportunity to intersect NP delivery metrics with additional datasets could add a new dimension to our existing findings.

One strength of our screening approach is the use of robust analytical tools, such as univariate analyses and random forest algorithms, which enabled us to identify biomarkers correlated with NP association. The robust and quantitative manner in which we detected EGFR hits for antibodies as well as antibody-targeted NPs shows the utility of this platform for the development and optimization of targeted drug delivery platforms, including antibody-targeted NPs, and its potential to apply to other targeted therapeutics, including ADCs. This method of analysis could provide therapeutic insights in the design of ADCs, specifically in evaluating the effects of conjugation site or linker chemistry.

By clustering NP-specific biomarkers across formulations, we constructed interaction networks, identifying and connecting genes associated with NP binding, recognition, and subcellular trafficking. This provides the scientific community with a blueprint for the fundamental study of cellular processes that mediate NP engagement, with applications for both basic and translational research.

We identified expression of SLC46A3, a lysosomal transporter, to be a negative regulator and potential biomarker for lipid-based NP uptake and downstream functional efficacy. Although SLC46A3 has recently been implicated in hepatic copper homeostasis as well as sensitivity to ADCs in cancer cells (*54*–*56*), its role in NP delivery was previously unexplored. We first validated SLC46A3 as a negative regulator of lipid-based NP uptake in a panel of nonpooled cell lines, as well as engineered isogenic cell lines with modulated *SLC46A3* expression. Because all current FDA-approved NPs for anticancer applications are liposomal formulations, there is notable potential for this biomarker to be quickly implemented in clinical studies with existing, approved formulations. To this end, we recapitulated our findings in an in vivo model using an analog of an FDA-approved liposomal NP formulation.

Moreover, we demonstrated that SLC46A3 has potential as a predictive biomarker beyond liposomal NPs by investigating solid lipid NPs. Both LNP-cell association and mRNA transfection were inversely correlated with SLC46A3 expression. These preliminary findings suggest that SLC46A3 expression may serve as a predictive biomarker for functional delivery of nucleic acid cargo through lipid NPs. Our findings support the continued exploration of SLC46A3 as a potential biomarker for therapeutic NP delivery.

We present a platform to study NP–cancer cell interactions simultaneously through the use of pooled screening, genomics, and machine learning algorithms. Application of this integrated platform should advance the rational design of nanocarriers.

Materials and methods

Extended materials and methods are available in the supplementary materials (*29*).

Base liposome synthesis

A thin film was generated from a lipid mixture composed of 31 mol % 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC), 31 mol % cholesterol, 31 mol % 1,2-distearoyl-sn-glycero-3-phospho-(1′-rac-glycerol) (DSPG), and 6 mol % 1,2-distearoyl-sn-glycero-3-phosphoethanolamine (DSPE) (Avanti) and rehydrated to 2 mg/ml under heat (65°C) and sonication. The liposome suspension was extruded by using an Avestin LiposoFast LF-50 liposome extruder to a diameter of 50 to 100 nm. Liposomes were fluorescently labeled through *N*-hydroxysuccinimide (NHS)–coupling of sulfo-cyanine NHS ester dye to DSPE headgroups according to the dye manufacturer (Lumiprobe) instructions. Lipid film generation, rehydration, extrusion, and dye labeling steps were similarly applied to all liposome formulations unless noted otherwise.

Tangential flow filtration (TFF)

To remove excess dye, crude NP solution was connected to a Spectrum Labs KrosFlo II system by using masterflex, Teflon-coated tubing. D02-E100-05-N membranes were used to purify the particles until dye was no longer seen in the permeate. Samples were run at flow rates of 80 ml/min with size 16 tubing. Phosphate-buffered saline (PBS) was used as the exchange buffer for the first five washes followed by milliQ water for the rest of the purification steps. After TFF, liposomes were characterized by means of dynamic light scattering (DLS). For layer-by-layer (LbL) synthesis, TFF was used for purification after deposition of each polyelectrolyte layer, following the above procedure. Instead of PBS, only milliQ water was passed through the TFF for LbL NP purification.

PLGA NP synthesis

PLGA (Sigma Aldrich) was dissolved at a concentration of 10 mg/ml in acetone, and Cy5 free acid dye (Lumiprobe) was dissolved at a concentration of 50 mg/ml in dimethyl sulfoxide (DMSO). 6 ml milliQ water were added to a scintillation vial and stirred gently on a stir plate; 2 μl dye were mixed with 1 ml PLGA solution and drawn up in a syringe with a 27-gauge needle attached. The PLGA-Cy5 solution was slowly added to the water under constant stirring and left to stir 3 hours. An additional 2 ml milliQ water were added the solution before purification by using TFF.

Synthesis of layer-by-layer NPs

Liposomes and PLGA NPs were layered by adding equal volumes of NP solution to polyelectrolyte solution under sonication. Polyelectrolyte solutions for liposome layering, with the exception of HA and alginate, were prepared in 50 mM hydroxyethylpiperazine ethane sulfonic acid (HEPES) and 40 mM NaCl (pH 7.4). HA and alginate stocks were prepared in 10 mM HEPES. All polyelectrolyte solutions for PLGA NP layering were prepared in water. Layered particles were incubated at room temperature for 1 hour before being purified by means of TFF and characterized by means of DLS.

Pooled PRISM cell dosing with NPs and preparation for flow cytometry

Cells were seeded at 200,000 cells/well in 0.5 ml RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS) in a 12-well plate. Cells were allowed to grow for 24 hours before treatment with NPs for 4 or 24 hours. After incubation, cells were washed once with warm PBS and dissociated with 0.25% Trypsin-EDTA. After 5 min at 37°C, the trypsin quenched with cell culture medium. Cells were then transferred to a FACS tube through a cell strainer cap and placed on ice until sorting.

Pooled PRISM cell dosing with antibodies

Cells were washed and dissociated with StemPro Accutase. After incubation, cold FACS buffer (PBS + 2% FBS) was added to each well, and cells were triturated and centrifuged. After spinning, cells were resuspended in FACS buffer at a concentration of 1×10^6 cells/ml. The cell solution was split into four groups: untreated control, (+) 15 μl 0.1 mg/ml Cy5-cetuximab, (+)15 μl 0.1 mg/ml Cy5-IgG, and (+) 5 μl of EGFR-AF488 (used at undiluted stock concentration provided by manufacturer, InvivoGen). Samples were incubated in the dark at 4°C for 1 hour. Cells were then washed and resuspended in cold FACS buffer.

SLC46A3 validation studies

Nonpooled screening

HCC1143 (RPMI-1640), HCC1395 (RPMI-1640), HeLa (RPMI-1640), SW948 (RPMI-1640), LOXIMVI (RPMI-1640), SJSA-1 (RPMI-1640), MCF7 [Eagle's minimum essential medium (EMEM)], DAOY (EMEM), MDA-MB-231 [Dulbecco's modified Eagle's medium (DMEM)], CAOV3 (DMEM), T47D (RPMI-1640), and HepG2 (DMEM) cells were seeded individually at 10,000 cells/well in 100 μl medium, supplemented with 10% FBS and 1X Penicillin-Streptomycin. Cells were allowed to grow overnight before treatment with NPs. Before dosing, all NP formulations were normalized to a concentration of 50 μg/ml. Cells were dosed with 10 μl normalized NP solutions. After incubation, cells were washed once with warm PBS and dissociated with 0.25% Trypsin-EDTA before quenching with cell culture medium. Cells were placed on ice until analyzed by using a high-throughput analyzer.

SLC46A3 overexpression: Viral transfection of LOXIMVI cells

Lentiviral vectors were purchased from the Broad Institute's Genetic Perturbation Platform (GPP), specifically ccsbBroad304_09945 (SLC46A3) and ccsbBroad304_99991 (Luciferase, vector control). LOXIMVI cells were trypsinized, counted, and resuspended to a concentration of 1.36×10^6 cells/ml. A solution of 2X polybrene was added to the cell suspension so that the final concentration of polybrene was 8 μg/ml. Cells were seeded into two six-well plates at 750,000 cells/well. Lentiviral vectors were separately added to plates at six different doses: 0, 25, 50, 100, 200, and 400 μl. After, 1 ml medium was added to each well, the cells were incubated overnight, and the medium was changed at 17 hours after seeding. At 48 hours after seeding, the cells were reseeded at 375,000 cells/well in 2 ml blasticidin containing medium (final blasticidin concentration was 1 μg/ml). The selection progress was monitored with flow cytometry (fig. S16).

SLC46A3 permanent deactivation with CRISPR-Cas9 in T47D cells

SLC46A3-deactivated T47D cell lines were generated through infection with lentiCRISPRv2-Opti (Addgene 163126) vectors encoding Cas9 and single-guide RNAs (sgRNAs) (*68*). The following oligonucleotides were used for sgRNA cloning and include cloning overhangs for ligation after BsmBI digest of lentiCRISPRv2-Opti vector: sgGFP_F, caccGGGCGAGGAGCTGTTCACCG; sgGFP_R, aaacCGGTGAACAGCTCCTCGCCC; sgSLC46A3_F, caccgAAAGCAAGCTCCCCAAAATG; and sgSLC46A3_R, aaacCATTTTGGGGAGCTTGCTTTc.

Clonal deactivation cell lines were isolated through FACS, and biallelic frame-shifts were confirmed with deep-sequencing [allele 1, −32 base pairs (bp) frameshift 501 reads; allele 2, −10 bp frameshift; 477 reads). The T47D *SLC46A3*-deactivation line described has the mutant alleles c.442_453del and c.440_449del.

Animal studies

All animal experiments were approved by the Massachusetts Institute of Technology Committee on Animal Care (CAC; protocol number 0821-052-04) and were conducted under the oversight of the Division of Comparative Medicine (DCM). Flank tumors of LOXIMVI-vector control and LOXIMVI-SLC46A3 OE cells were established with a subcutaneous injection of 0.5×10^6 to 1.0×10^6 cells as a 1:1 mixture with MatriGel (Corning) and PBS to the right flank of NCr nude mice (5 to 7 weeks, Taconic, NCRNU-F, NCRNU-M). Sample sizes of studies were initially determined by use of G*Power.

For intratumoral studies, mice with established flank tumors were randomly assigned to either the 4- or 24-hour dosing cohort. After injection, mice were imaged by using the In Vivo Imaging System (IVIS) Spectrum whole-animal imaging device (PerkinElmer) using ex = 640/em = 700 nm to capture Cy5 signal. Immediately after imaging, mice were humanely euthanized, and tumors were excised and imaged again with IVIS. For intravenous studies, NPs were administered to mice by using tail vein injections, each of three doses spaced 24 hours apart. Four hours after the third and final injection, mice were humanely euthanized, and tumors were excised and imaged with IVIS. Tumors were weighed, and their weights were recorded for normalization of tumor fluorescence by tumor mass.

Statistical analysis

All statistical analysis for nonpooled validation studies was performed by using GraphPad PRISM 9. Detailed statistical information is provided for each figure in the associated caption. Unless noted otherwise, for single comparisons (nonparametric), the Mann-Whitney test was used. For multiple comparison testing, the Kruskall-Wallis test was used to compare treatment groups with the parental control. The datasets and code pertaining to nanoPRISM probabilistic model development and subsequent analyses (univariate and random forest) are available on Zenodo (*69*–*75*).

REFERENCES AND NOTES

1. J. Shi, P. W. Kantoff, R. Wooster, O. C. Farokhzad, Cancer nanomedicine: Progress, challenges and opportunities. *Nat. Rev. Cancer* **17**, 20–37 (2017). doi: 10.1038/nrc.2016.108; pmid: 27834398
2. M. J. Mitchell *et al*., Engineering precision nanoparticles for drug delivery. *Nat. Rev. Drug Discov.* **20**, 101–124 (2021). doi: 10.1038/s41573-020-0090-8; pmid: 33277608
3. N. Boehnke, P. T. Hammond, Power in numbers: Harnessing combinatorial and integrated screens to advance nanomedicine. *JACS Au* **2**, 12–21 (2021). doi: 10.1021/jacsau.1c00313; pmid: 35098219
4. S. Tran, P. J. DeGiovanni, B. Piel, P. Rai, Cancer nanomedicine: A review of recent success in drug delivery. *Clin. Transl. Med.* **6**, 44 (2017). doi: 10.1186/s40169-017-0175-0; pmid: 29230567
5. S. Wilhelm *et al*., Analysis of nanoparticle delivery to tumours. *Nat. Rev. Mater.* **1**, 16014 (2016). doi: 10.1038/natrevmats.2016.14
6. Y. H. Cheng, C. He, J. E. Riviere, N. A. Monteiro-Riviere, Z. Lin, Meta-analysis of nanoparticle delivery to tumors using a physiologically based pharmacokinetic modeling and simulation approach. *ACS Nano* **14**, 3075–3095 (2020). doi: 10.1021/acsnano.9b08142; pmid: 32078303
7. Y. S. Youn, Y. H. Bae, Perspectives on the past, present, and future of cancer nanomedicine. *Adv. Drug Deliv. Rev.* **130**, 3–11 (2018). doi: 10.1016/j.addr.2018.05.008; pmid: 29778902
8. W. Poon, B. R. Kingston, B. Ouyang, W. Ngo, W. C. W. Chan, A framework for designing delivery systems. *Nat. Nanotechnol.* **15**, 819–829 (2020). doi: 10.1038/s41565-020-0759-5; pmid: 32895522
9. W. Poon *et al*., Elimination pathways of nanoparticles. *ACS Nano* **13**, 5785–5798 (2019). doi: 10.1021/acsnano.9b01383; pmid: 30990673
10. S. Correa *et al*., Tuning nanoparticle interactions with ovarian cancer through layer-by-layer modification of surface chemistry. *ACS Nano* **14**, 2224–2237 (2020). doi: 10.1021/acsnano.9b09213; pmid: 31971772

11. N. Boehnke *et al*., Theranostic layer-by-layer nanoparticles for simultaneous tumor detection and gene silencing. *Angew. Chem. Int. Ed.* **59**, 2776–2783 (2020). doi: 10.1002/anie.201911762; pmid: 31747099
12. N. Boehnke, K. J. Dolph, V. M. Juarez, J. M. Lanoha, P. T. Hammond, Electrostatic conjugation of nanoparticle surfaces with functional peptide motifs. *Bioconjug. Chem.* **31**, 2211–2219 (2020). doi: 10.1021/acs.bioconjchem.0c00384; pmid: 32786506
13. J. E. Dahlman *et al*., Barcoded nanoparticles for high throughput in vivo discovery of targeted therapeutics. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 2060–2065 (2017). doi: 10.1073/pnas.1620874114; pmid: 28167778
14. B. Nogrady, How cancer genomics is transforming diagnosis and treatment. *Nature* **579**, S10–S11 (2020). doi: 10.1038/d41586-020-00845-4; pmid: 32214255
15. C. Yu *et al*., High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. *Nat. Biotechnol.* **34**, 419–423 (2016). doi: 10.1038/nbt.3460; pmid: 26928769
16. S. M. Corsello *et al*., Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. *Nature Cancer*, (2020).
17. S. Correa, N. Boehnke, E. Deiss-Yehiely, P. T. Hammond, Solution conditions tune and optimize loading of therapeutic polyelectrolytes into layer-by-layer functionalized liposomes. *ACS Nano* **13**, 5623–5634 (2019). doi: 10.1021/acsnano.9b00792; pmid: 30986034
18. S. Correa *et al*., Highly scalable, closed-loop synthesis of drug-loaded, layer-by-layer nanoparticles. *Adv. Funct. Mater.* **26**, 991–1003 (2016). doi: 10.1002/adfm.201504385; pmid: 27134622
19. Z. J. Deng *et al*., Layer-by-layer nanoparticles for systemic codelivery of an anticancer drug and siRNA for potential triple-negative breast cancer treatment. *ACS Nano* **7**, 9571–9584 (2013). doi: 10.1021/nn4047925; pmid: 24144228
20. S. W. Morton, Z. Poon, P. T. Hammond, The architecture and biological performance of drug-loaded LbL nanoparticles. *Biomaterials* **34**, 5328–5335 (2013). doi: 10.1016/j.biomaterials.2013.03.059; pmid: 23618629
21. G. Decher, Fuzzy nanoassemblies: Toward layered polymeric multicomposites. *Science* **277**, 1232–1237 (1997). doi: 10.1126/science.277.5330.1232
22. E. C. Dreaden *et al*., Tumor-targeted synergistic blockade of MAPK and PI3K from a layer-by-layer nanoparticle. *Clin. Cancer Res.* **21**, 4410–4419 (2015). doi: 10.1158/1078-0432.CCR-15-0013; pmid: 26034127
23. E. C. Dreaden *et al*., Bimodal tumor-targeting from microenvironment responsive hyaluronan layer-by-layer (LbL) nanoparticles. *ACS Nano* **8**, 8374–8382 (2014). doi: 10.1021/nn502861t; pmid: 25100313
24. O. P. Oommen, C. Duehrkop, B. Nilsson, J. Hilborn, O. P. Varghese, Multifunctional hyaluronic acid and chondroitin sulfate nanoparticles: Impact of glycosaminoglycan presentation on receptor mediated cellular uptake and immune activation. *ACS Appl. Mater. Interfaces* **8**, 20614–20624 (2016). doi: 10.1021/acsami.6b06823; pmid: 27468113
25. J. S. Suk, Q. Xu, N. Kim, J. Hanes, L. M. Ensign, PEGylation as a strategy for improving nanoparticle-based drug and gene delivery. *Adv. Drug Deliv. Rev.* **99** (Pt A), 28–51 (2016). doi: 10.1016/j.addr.2015.09.012; pmid: 26456916
26. E. Fröhlich, The role of surface charge in cellular uptake and cytotoxicity of medical nanoparticles. *Int. J. Nanomedicine* **7**, 5577–5591 (2012). doi: 10.2147/IJN.S36111; pmid: 23144561
27. E. A. Berg, J. B. Fishman, Labeling antibodies using *N*-hydroxysuccinimide (NHS)–fluorescein. *Cold Spring Harb. Protoc.* **2019**, 229–231 (2019). pmid: 30824621
28. X. Jin *et al*., A metastasis map of human cancer cell lines. *Nature* **588**, 331–336 (2020). doi: 10.1038/s41586-020-2969-2; pmid: 33299191
29. Materials and methods are available as supplementary materials.
30. J. A. Kim, C. Åberg, A. Salvati, K. A. Dawson, Role of cell cycle on the cellular uptake and dilution of nanoparticles in a cell population. *Nat. Nanotechnol.* **7**, 62–68 (2011). doi: 10.1038/nnano.2011.191; pmid: 22056728
31. C. Åberg, J. A. Kim, A. Salvati, K. A. Dawson, Reply to 'The interface of nanoparticles with proliferating mammalian cells'. *Nat. Nanotechnol.* **12**, 600–603 (2017). doi: 10.1038/nnano.2017.139; pmid: 28681851
32. E. Panet *et al*., The interface of nanoparticles with proliferating mammalian cells. *Nat. Nanotechnol.* **12**, 598–600 (2017). doi: 10.1038/nnano.2017.140; pmid: 28681852
33. P. Rees, J. W. Wills, M. R. Brown, C. M. Barnes, H. D. Summers, The origin of heterogeneous nanoparticle uptake by cells. *Nat. Commun.* **10**, 2341 (2019). doi: 10.1038/s41467-019-10112-4; pmid: 31138801
34. J. Barretina *et al*., The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012). doi: 10.1038/nature11735; pmid: 22460905
35. M. Ghandi *et al*., Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019). doi: 10.1038/s41586-019-1186-3; pmid: 31068700
36. K. Tsuchikama, Z. An, Antibody-drug conjugates: Recent advances in conjugation and linker chemistries. *Protein Cell* **9**, 33–46 (2018). doi: 10.1007/s13238-016-0323-0; pmid: 27743348
37. J. Rejman, V. Oberle, I. S. Zuhorn, D. Hoekstra, Size-dependent internalization of particles via the pathways of clathrin- and caveolae-mediated endocytosis. *Biochem. J.* **377**, 159–169 (2004). doi: 10.1042/bj20031253; pmid: 14505488
38. S. Behzadi *et al*., Cellular uptake of nanoparticles: Journey inside the cell. *Chem. Soc. Rev.* **46**, 4218–4244 (2017). doi: 10.1039/C6CS00636A; pmid: 28585944
39. A. Subramanian *et al*., Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005). doi: 10.1073/pnas.0506580102; pmid: 16199517
40. A. Liberzon *et al*., Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011). doi: 10.1093/bioinformatics/btr260; pmid: 21546393
41. A. Liberzon *et al*., The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015). doi: 10.1016/j.cels.2015.12.004; pmid: 26771021
42. M. Dean, Y. Hamon, G. Chimini, The human ATP-binding cassette (ABC) transporter superfamily. *J. Lipid Res.* **42**, 1007–1017 (2001). doi: 10.1016/S0022-2275(20)31588-1; pmid: 11441126
43. Y. Shamay *et al*., P-selectin is a nanotherapeutic delivery target in the tumor microenvironment. *Sci. Transl. Med.* **10**, 345ra87 (2018). pmid: 27358497
44. G. Saravanakumar, D. G. Jo, J. H. Park, Polysaccharide-based nanoparticles: A versatile platform for drug delivery and biomedical imaging. *Curr. Med. Chem.* **19**, 3212–3229 (2012). doi: 10.2174/092986712800784658; pmid: 22612705
45. J. Voigt, J. Christensen, V. P. Shastri, Differential uptake of nanoparticles by endothelial cells through polyelectrolytes with affinity for caveolae. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2942–2947 (2014). doi: 10.1073/pnas.1322356111; pmid: 24516167
46. D. Szklarczyk *et al*., STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019). doi: 10.1093/nar/gky1131; pmid: 30476243
47. C. von Mering *et al*., STRING: A database of predicted functional associations between proteins. *Nucleic Acids Res.* **31**, 258–261 (2003). doi: 10.1093/nar/gkg034; pmid: 12519996
48. B. Snel, G. Lehmann, P. Bork, M. A. Huynen, STRING: A web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* **28**, 3442–3444 (2000). doi: 10.1093/nar/28.18.3442; pmid: 10982861
49. D. Martin *et al*., GOToolBox: Functional analysis of gene datasets based on Gene Ontology. *Genome Biol.* **5**, R101 (2004). doi: 10.1186/gb-2004-5-12-r101; pmid: 15575967
50. M. Ashburner *et al*., Gene Ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000). doi: 10.1038/75556; pmid: 10802651
51. S. Carbon *et al*., The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Res.* **49** (D1), D325–D334 (2021). doi: 10.1093/nar/gkaa1113; pmid: 33290552
52. L. Lin, S. W. Yee, R. B. Kim, K. M. Giacomini, SLC transporters as therapeutic targets: Emerging opportunities. *Nat. Rev. Drug Discov.* **14**, 543–560 (2015). doi: 10.1038/nrd4626; pmid: 26111766
53. A. Chapel *et al*., An extended proteome map of the lysosomal membrane reveals novel potential transporters. *Mol. Cell. Proteomics* **12**, 1572–1588 (2013). doi: 10.1074/mcp.M112.021980; pmid: 23436907
54. J. H. Kim *et al*., Lysosomal SLC46A3 modulates hepatic cytosolic copper homeostasis. *Nat. Commun.* **12**, 290 (2021). doi: 10.1038/s41467-020-20461-0; pmid: 33436590
55. K. J. Hamblett *et al*., SLC46A3 is required to transport catabolites of noncleavable antibody maytansine conjugates from the lysosome to the cytoplasm. *Cancer Res.* **75**, 5329–5340 (2015). doi: 10.1158/0008-5472.CAN-15-1610; pmid: 26631267
56. K. Kinneer *et al*., SLC46A3 as a potential predictive biomarker for antibody-drug conjugates bearing noncleavable linked maytansinoid and pyrrolobenzodiazepine warheads. *Clin. Cancer Res.* **24**, 6570–6582 (2018). doi: 10.1158/1078-0432.CCR-18-1300; pmid: 30131388
57. Q. Zhao *et al*., Increased expression of SLC46A3 to oppose the progression of hepatocellular carcinoma and its effect on sorafenib therapy. *Biomed. Pharmacother.* **114**, 108864 (2019). doi: 10.1016/j.biopha.2019.108864; pmid: 30981107
58. G. Li *et al*., Mechanisms of acquired resistance to trastuzumab emtansine in breast cancer cells. *Mol. Cancer Ther.* **17**, 1441–1453 (2018). doi: 10.1158/1535-7163.MCT-17-0296; pmid: 29695635
59. C. K. Tsui *et al*., CRISPR-Cas9 screens identify regulators of antibody-drug conjugate toxicity. *Nat. Chem. Biol.* **15**, 949–958 (2019). doi: 10.1038/s41589-019-0342-2; pmid: 31451760
60. X. Hou, T. Zaks, R. Langer, Y. Dong, Lipid nanoparticles for mRNA delivery. *Nat. Rev. Mater.* **6**, 1078–1094 (2021). doi: 10.1038/s41578-021-00358-0; pmid: 34394960
61. E. Samaridou, J. Heyes, P. Lutwyche, Lipid nanoparticles for nucleic acid delivery: Current perspectives. *Adv. Drug Deliv. Rev.* **154-155**, 37–63 (2020). doi: 10.1016/j.addr.2020.06.002; pmid: 32526452
62. M. Jayaraman *et al*., Maximizing the potency of siRNA lipid nanoparticles for hepatic gene silencing in vivo. *Angew. Chem. Int. Ed.* **51**, 8529–8533 (2012). doi: 10.1002/anie.201203263; pmid: 22782619
63. K. A. Whitehead *et al*., Degradable lipid nanoparticles with predictable in vivo siRNA delivery activity. *Nat. Commun.* **5**, 4277 (2014). doi: 10.1038/ncomms5277; pmid: 24969323
64. K. J. Kauffman *et al*., Optimization of lipid nanoparticle formulations for mRNA delivery in vivo with fractional factorial and definitive screening designs. *Nano Lett.* **15**, 7300–7306 (2015). doi: 10.1021/acs.nanolett.5b02497; pmid: 26469188
65. K. J. Hassett *et al*., Optimization of lipid nanoparticles for intramuscular administration of mRNA vaccines. *Mol. Ther. Nucleic Acids* **15**, 1–11 (2019). doi: 10.1016/j.omtn.2019.01.013; pmid: 30785039
66. Q. Q. Liu *et al*., Inorganic nanoparticles applied as functional therapeutics. *Adv. Funct. Mater.* **31**, 2008171 (2021).
67. W. Paul, C. P. Sharma, Inorganic nanoparticles for targeted drug delivery. *Biointegration Med. Implant. Mater.* **2020**, 334–373 (2020).
68. C. H. Adelmann *et al*., MFSD12 mediates the import of cysteine into melanosomes and lysosomes. *Nature* **588**, 699–704 (2020). doi: 10.1038/s41586-020-2937-x; pmid: 33208952
69. N. Boehnke, J. P. Straehla, M. Kocak, M. Ronan, J. Roth, Data and processing scripts for PRISM barcode sequencing data used in "Massively parallel pooled screening reveals genomic determinants of nanoparticle-cell interactions". *Zenodo* (2022); doi: 10.5281/zenodo.6642633
70. M. Stephens, False discovery rates: A new deal. *Biostatistics* **18**, 275–294 (2017). pmid: 27756721
71. R Core Team, R: A language and envrionemnt for statistical computing. R Foundation for Statistical Compouting; https://www.R-project.org.
72. H. Wickham, ggplot2: Elegant graphics for data analysis. *Use R*, 1-212 (2009).
73. M. Kocak, A. Boghossian, Linear association function used in the manuscript "Massively parallel pooled screening reveals genomic determinants of nanoparticle-cell interactions". *Zenodo* (2022); doi: 10.5281/zenodo.6558445
74. Y. Tang, M. Horikoshi, W. X. Li, ggfortify: Unified interface to visualize statistical results of popular R packages. *R J.* **8**, 474–485 (2016). doi: 10.32614/RJ-2016-060
75. M. Horikoshi, Y. Tang, ggfortify: Data visualization tools for statistical analysis results(2016); https://CRAN.R-project.org/package=ggfortify.

ACKNOWLEDGMENTS

We thank the Koch Institute's Robert A. Swanson (1969) Biotechnology Center for technical support, specifically the Flow Cytometry, High Throughput Sciences, Genomics Core, Microscopy, and Preclinical Modeling, Imaging and Testing cores; the Hope Babette Tang (1983) Histology Facility; and the Peterson (1957) Nanotechnology Materials Core Facility. We thank T. Golub and A. Burgin for formative feedback and helpful discussion. We also gratefully acknowledge T. Diefenbach and the Ragon Institute Microscopy Core for assistance with imaging cytometry. The

LOXIMVI and T47D cell lines were gifts from the F. Gertler laboratory (MIT), and the Jurkat cell line was a gift from the D. Sabatini laboratory (previously of the Whitehead Institute for Biomedical Research). We thank C. Straehla for help in figure design. Figures 1A and 6A were created in part by use of Biorender.com. **Funding:** This work was supported in part by SPARC funding at The Broad Institute. This work was also supported by grants from the Koch Institute's Marble Center for Cancer Nanomedicine and Frontier Research Program. This work was supported in part by the Koch Institute Support (core) Grant P30-CA14051 from the National Cancer Institute. N.B. was supported by a US Department of Defense Congressionally Directed Medical Research Programs Peer Reviewed Cancer Research Program Horizon Award (W81XWH-19-1-0257) and the NIH-NCI (K99CA255844). J.P.S. was supported as a National Institutes of Health (NIH) grant T32 trainee (CA136432-08) and by the Helen Gurley Brown Presidential Initiative of Dana-Farber Cancer Institute. Fellowship support for C.H.A. was from the NIH (NRSA F31 CA228241-01). R.R.C. is a fellow of the Parker B. Francis Foundation. N.N. was supported by a grant from the Gates Foundation. Fellowship support for A.G.B. was from the NIH (F30 DK130564) and a Termeer Fellowship of Medical Engineering and Science. N.G.L. was supported by Cancer Research UK and the Brain Tumour Charity (C42454/A28596) and a fellowship from the Ludwig Center at the Koch Institute for Integrative Cancer Research. H.L. was supported by the Charles W. (1955) and Jennifer C. Johnson Cancer Research Fund and NIH (K08DK123414). **Author contributions:** Conceptualization: N.B. and J.P.S. Methodology: N.B., J.P.S., and M.K. Formal Analysis: N.B., J.P.S., M.K., M.G.R., and M.R. Investigation: N.B., J.P.S., H.C.S., M.G.R., D.R., N.N., A.G.B., and N.G.L. Visualization: N.B. and J.P.S. Funding acquisition: N.B., J.P.S., A.N.K., and P.T.H. Project administration: N.B., J.P.S., and M.R. Validation: N.B., J.P.S., H.C.S., C.H.A., R.R.C., J.H.C., and H.L. Supervision: J.A.R., A.N.K., and P.T.H. Writing – original draft: N.B. and J.P.S. Writing – review and editing: N.B., J.P.S., H.C.S., M.K., M.G.R., M.R., C.H.A., R.R.C., N.N., A.G.B., N.G.L., J.H.C., H.L., J.A.R., A.N.K., and P.T.H. **Competing interests:** N.B., J.P.S., A.N.K., and P.T.H. have submitted a patent application for the SLC46A3 biomarker discovery. A.N.K. is a founder and member of the board of 76Bio. P.T.H. is a member of the science advisory board at Moderna; board member at Alector, Advanced Chemotherapy Technologies, and Burroughs-Wellcome Fund; and board member and cofounder of LayerBio. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

Submitted 24 September 2021; resubmitted 14 March 2022
Accepted 16 June 2022
10.1126/science.abm5551

RESEARCH ARTICLE SUMMARY

CORONAVIRUS

Pathogen-sugar interactions revealed by universal saturation transfer analysis

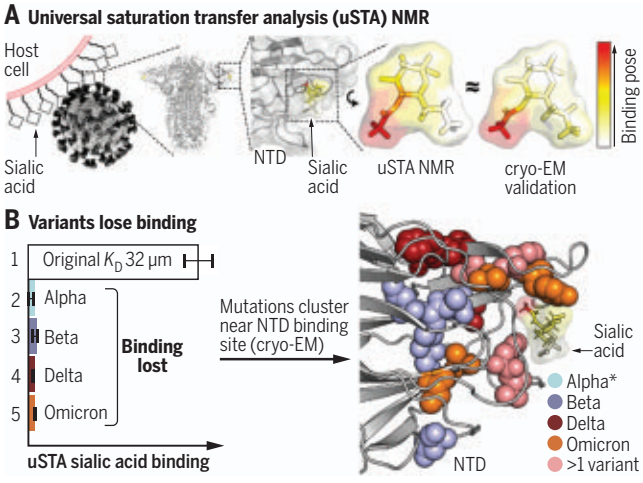
Charles J. Buchanan†, Ben Gaunt†, Peter J. Harrison, Yun Yang, Jiwei Liu, Aziz Khan, Andrew M. Giltrap, Audrey Le Bas, Philip N. Ward, Kapil Gupta, Maud Dumoux, Tiong Kit Tan, Lisa Schimaski, Sergio Daga, Nicola Picchiotti, Margherita Baldassarri, Elisa Benetti, Chiara Fallerini, Francesca Fava, Annarita Giliberti, Panagiotis I. Koukos, Matthew J. Davy, Abirami Lakshminarayanan, Xiaochao Xue, Georgios Papadakis, Lachlan P. Deimel, Virgínia Casablancas-Antràs, Timothy D. W. Claridge, Alexandre M. J. J. Bonvin, Quentin J. Sattentau, Simone Furini, Marco Gori, Jiandong Huo, Raymond J. Owens, Christiane Schaffitzel, Imre Berger, Alessandra Renieri, GEN-COVID Multicenter Study, James H. Naismith*, Andrew J. Baldwin*, Benjamin G. Davis*

INTRODUCTION: The surface proteins found on both pathogens and host cells mediate cell entry (and exit) and influence disease progression and transmission. Both types of proteins can bear host-generated posttranslational modifications, such as glycosylation, that are essential for function but can confound current biophysical methods used for dissecting key interactions. Several human viruses (including non-SARS coronaviruses) attach to host cell surface N-linked glycans that include forms of sialic acid (sialosides). There remains, however, conflicting evidence as to whether or how SARS-associated coronaviruses might use such a mechanism. In the absence of an appropriate biochemical assay, the ability to analyze the binding of such glycans to heavily modified proteins and resolve this issue is limited.

RATIONALE: We developed and demonstrated a quantitative extension of "saturation transfer" protein nuclear magnetic resonance (NMR) methods to a complete mathematical model of the magnetization transfer caused by interactions between protein and ligand. The designed method couples objective resonance identification and intensity measurement in NMR spectra (via a deconvolution algorithm) with Bloch-McConnell analysis of magnetization transfer (as judged by this resonance signal intensity) to enable a structural, kinetic, and thermodynamic analysis of ligand binding. Such quantification is beyond previously perceived limits of exchange rates, concentration, or system and therefore represents a potentially universal saturation transfer analysis (uSTA) method.

RESULTS: In an automated workflow, uSTA can be applied to a range of even heavily modified protein systems in a general manner to obtain quantitative binding interaction parameters (K_D, k_{Ex}). uSTA proved critical in mapping direct interactions between sialoside sugar ligands and relevant virus surface attachment glycoproteins, including multiple variants of both severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein and influenza H1N1 hemagglutinin protein. It was successful in quantitating ligand NMR signals in spectral regions otherwise occluded by resonances from mobile protein glycans. In early-pandemic (December 2019) B-origin-lineage SARS-CoV-2 spike trimer, a clear "end-on" binding mode of sialoside sugars to spike was revealed by uSTA. This mode contrasted with "extended-surface side"-binding for heparin sugar ligands. uSTA-derived restraints used in structural modeling suggested sialoside-glycan binding sites in a β sheet–rich region of spike N-terminal domain (NTD), distant from the receptor-binding domain (RBD) that binds ACE2 co-receptor and that has been identified as the site for other sugar interactions. Consistent with this NTD site being a previously unknown sialoside sugar-binding pocket, uSTA-sialoside binding was minimally perturbed by antibodies that neutralize the ACE2-binding RBD domain. Strikingly, uSTA also shows that this sialoside binding is disrupted in spike from multiple variants of concern (B1.1.7/alpha, B1.351/beta, B.1.617.2/delta, and B.1.1.529/omicron) that emerged later in the pandemic (September 2020 onward). Notably, these variants possess multiple hotspot mutations in the NTD. End-on sialoside binding in a B-origin-lineage spike-NTD pocket was pinpointed by cryo-EM to a previously unknown site that is created from residues that are notably mutated or are in regions where mutations occur in variants of concern (e.g., His^{69}, Val^{70}, and Tyr^{145} in alpha and omicron). An analysis of beneficial genetic variances correlated with disease severity in cohorts of patients from early 2020 suggests a model in which this site in the NTD of B-origin-lineage SARS-CoV-2 (but not in later variants) may have exploited a specific sialylated polylactosamine motif found on tetra-antennary human N-linked glycoproteins, known to be present in deeper human lung.

CONCLUSION: Together, these results confirm a distinctive sugar-binding mode mediated by the unusual NTD of B-origin-lineage SARS-CoV-2 spike protein that is lost in later variants. This may implicate modulation of binding by SARS-CoV-2 virus to human cell surface sugars as a determinant of virulence and/or zoonosis. More generally, because cell surface glycans are widely relevant to biology and pathology, the uSTA method can now provide ready, quantitative, widespread analysis of complex, host-derived, and posttranslationally modified proteins in their binding to putative ligands, which may be relevant to disease, even in previously confounding complex systems. ■



uSTA identifies a sugar-binding site and pose in the unusual NTD of early-pandemic SARS-CoV-2 spike. (**A**) This site and pose, which were confirmed by cryo–electron microscopy, are heavily mutated in variants of concern. (**B**) Analyses reveal a loss of sialoside-spike binding, rationalized by clustering of mutations around the binding site. *All alpha-variant mutations are in >1 variant.

The list of author affiliations is available in the full article online.
***Corresponding author. Email: naismith@strubi.ox.ac.uk (J.H.N.); andrew.baldwin@chem.ox.ac.uk (A.J.B.); ben.davis@rfi.ac.uk (B.G.D.)**
†These authors contributed equally to this work.

Cite this article as C. J. Buchanan *et al.*, *Science* 377, eabm3125 (2022). DOI: 10.1126/science.abm3125

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abm3125

RESEARCH ARTICLE

CORONAVIRUS

Pathogen-sugar interactions revealed by universal saturation transfer analysis

Charles J. Buchanan[1,2,3]†, Ben Gaunt[1]†, Peter J. Harrison[4,5], Yun Yang[1,4], Jiwei Liu[1], Aziz Khan[1,2], Andrew M. Giltrap[1,2], Audrey Le Bas[1,4], Philip N. Ward[4], Kapil Gupta[6], Maud Dumoux[1], Tiong Kit Tan[7], Lisa Schimaski[7], Sergio Daga[8,9], Nicola Picchiotti[10,11], Margherita Baldassarri[8,9], Elisa Benetti[9], Chiara Fallerini[8,9], Francesca Fava[8,9,12], Annarita Giliberti[8,9], Panagiotis I. Koukos[13], Matthew J. Davy[1], Abirami Lakshminarayanan[1,2], Xiaochao Xue[2,14], Georgios Papadakis[2], Lachlan P. Deimel[14], Virgínia Casablancas-Antràs[2,3], Timothy D. W. Claridge[2], Alexandre M. J. J. Bonvin[13], Quentin J. Sattentau[14], Simone Furini[9], Marco Gori[10,15], Jiandong Huo[1,4], Raymond J. Owens[1,4], Christiane Schaffitzel[13], Imre Berger[13], Alessandra Renieri[8,9,12], GEN-COVID Multicenter Study, James H. Naismith[1,4]*, Andrew J. Baldwin[1,2,3]*, Benjamin G. Davis[1,2,16]*

Many pathogens exploit host cell-surface glycans. However, precise analyses of glycan ligands binding with heavily modified pathogen proteins can be confounded by overlapping sugar signals and/or compounded with known experimental constraints. Universal saturation transfer analysis (uSTA) builds on existing nuclear magnetic resonance spectroscopy to provide an automated workflow for quantitating protein-ligand interactions. uSTA reveals that early-pandemic, B-origin-lineage severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike trimer binds sialoside sugars in an "end-on" manner. uSTA-guided modeling and a high-resolution cryo–electron microscopy structure implicate the spike N-terminal domain (NTD) and confirm end-on binding. This finding rationalizes the effect of NTD mutations that abolish sugar binding in SARS-CoV-2 variants of concern. Together with genetic variance analyses in early pandemic patient cohorts, this binding implicates a sialylated polylactosamine motif found on tetraantennary N-linked glycoproteins deep in the human lung as potentially relevant to virulence and/or zoonosis.

Sialosides are present in glycans that are anchored to human cells, and they mediate binding that is central to cell-cell communication in human physiology and that is at the heart of many host-pathogen interactions. One of the most well-known examples is that of influenza virus, which binds to sialosides with its hemagglutinin (HA or H) protein and cleaves off sialic acid from the infected cell with its neuraminidase (NA or N) protein; HxNx variants of influenza with different HA or NA protein types have a profound effect on zoonosis and pathogenicity (*1*).

The Middle East respiratory syndrome [MERS (*2*)] virus, which is related to severe acute respiratory syndrome coronavirus 1 and 2 (SARS-CoV-1 and -2), has been shown to exploit cell-surface sugar sialosides (*2*–*6*) as part of an attachment strategy. Both SARS-CoV-1 (*7*–*9*) and SARS-CoV-2 (*10*, *11*) are known to gain entry to host cells through the use of receptor-binding domains (RBDs) of their respective spike proteins that bind human cell-surface protein ACE2, but whether these viruses engage sialosides as part of the infection cycle has, despite predictions (*6*, *12*), remained unclear. Preliminary reports as to whether complex sialosides are or are not bound are contradictory and format-dependent (*13*–*15*). Glycosaminoglycans on proteoglycans such as heparin have been identified as a primary cooperative glycan attachment point (*16*, *17*). Studies reporting sugar binding have so far implicated binding sites in or close to the RBD of the spike protein. Surprisingly, the N-terminal domain (NTD, fig. S1), which has a putative glycan binding fold (*10*, *18*, *19*) and binds sialosides in other non-SARS coronaviruses (including MERS), has been less explored. The NTD has no confirmed function in SARS-CoV-2, and yet neutralization by antibodies against this domain suggests a potentially important function in viral replication. The unresolved role of host cell-surface sialosides for this pathogen has been noted as an important open question (*17*). The hypothesized roles for sugar interactions (*20*) in both virulence (*21*) and zoonosis (*1*) indicate that there is an urgent need for precise, quantitative, and robust methods for analysis.

In principle, magnetization transfer in protein nuclear magnetic resonance (NMR) spectroscopy could meet this need, as it can measure ligand binding in its native state without the need for additional labeling or modification of either ligand or protein (e.g., attachment to surface or sensor) (figs. S2 and S3; see text S1 for more details). Saturation transfer difference (STD) (*22*), which has been widely used to gauge qualitative ligand-protein interactions (*23*), detects the transfer of magnetization while they are bound via "cross-relaxation."

In reality, complex, highly modified protein systems have proven difficult to analyze in a quantitative manner with current methods for several reasons. First, mammalian proteins (or those derived by pathogens from expression in infected mammalian hosts) often bear large, highly mobile glycans. Critically, in the case of glycoproteins such as SARS-CoV-2 spike that may themselves bind glycans, this leads to contributions to protein NMR spectra that may overlap with putative glycan ligand resonances, thereby obscuring needed signal. Second, the NMR spectra of glycan ligands are themselves complex, comprising many overlapped resonances as multiplets and limiting the accurate determination of signal intensities. Finally, STD is commonly described as limited to specific kinetic regimes and/or ligand-to-protein binding equilibrium positions (*24*). As a result, many regimes and systems have been considered inaccessible to STD.

Using a rigorous theoretical description, coupled with a computational approach based on a Bayesian deconvolution algorithm to objectively and accurately extract signal from all observed resonances, we have undertaken an optimized reformulation of the magnetization/"saturation" transfer protocol (figs. S4, S5, S6, and S8). This approach reliably and quantitatively determines precise binding rates (k_{on}, k_{off}, k_{ex}), constants (K_D), and interaction "maps" across a wide range of regimes (fig. S4), including systems previously thought to be intractable.

Design of uSTA based on a comprehensive treatment of ligand-protein magnetization transfer

While using existing STD methodology to study the interaction between the SARS-CoV-2 spike protein and sialosides, we noted several challenges that resulted in the development of

[1]Rosalind Franklin Institute, Harwell Science and Innovation Campus, Oxford OX11 0FA, UK. [2]Department of Chemistry, University of Oxford, Oxford OX1 3TA, UK. [3]Kavli Institute of Nanoscience Discovery, University of Oxford, Oxford OX1 3QU, UK. [4]Division of Structural Biology, Wellcome Centre for Human Genetics, University of Oxford, Headington, Oxford OX3 7BN, UK. [5]Diamond Light Source, Harwell Science and Innovation Campus, Oxfordshire, UK. [6]Max Planck Bristol Centre for Minimal Biology, University of Bristol, Bristol, UK. [7]MRC Human Immunology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK. [8]Medical Genetics, University of Siena, Siena, Italy. [9]Med Biotech Hub and Competence Center, Department of Medical Biotechnologies, University of Siena, Siena, Italy. [10]Department of Information Engineering and Mathematics, University of Siena, Siena, Italy. [11]Department of Mathematics, University of Pavia, Pavia, Italy. [12]Genetica Medica, Azienda Ospedaliera Universitaria Senese, Siena, Italy. [13]Bijvoet Centre for Biomolecular Research, Faculty of Science, Utrecht University, Utrecht, Netherlands. [14]Sir William Dunn School of Pathology, Oxford, UK. [15]Maasai, I3S CNRS, Université Côte d'Azur, Nice, France. [16]Department of Pharmacology, University of Oxford, Oxford OX1 3QT, UK.
*Corresponding author. Email: naismith@strubi.ox.ac.uk (J.H.N.); andrew.baldwin@chem.ox.ac.uk (A.J.B.); ben.davis@rfi.ac.uk (B.G.D.)
†These authors contributed equally to this work.

uSTA (see text S3 for more details). Our theoretical analyses (fig. S4) suggested that many common assumptions or limits that are thought to govern the applicability of magnetization transfer might in fact be circumvented, and we set out to devise a complete treatment that might accomplish this (figs. S5, S6, and S8). This resulted in five specific methodological changes that resulted in a more sensitive, accurate, quantitative, and general method for studying the interactions between biomolecules and ligands (summarized in figs. S5 and S8 and discussed in detail in text S4).

1) We noted the discrepancy between K_Ds determined using existing STD methods and those obtained using other biophysical methods (*25*). We performed a theoretical analysis using the Bloch-McConnell equations, a rigorous formulation for studying the evolution of magnetization in exchanging systems that has been widely used to analyze chemical exchange saturation transfer (*26*, *27*), dark-state exchange saturation transfer (*28*), and Carr-Purcell-Meiboom-Gill (*29*) NMR data to describe protein motion. This analysis not only allowed us to explain this discrepancy, but also enabled fitting of data to give accurate k_{on}, k_{off}, and K_D values for protein-ligand interactions that were in excellent accord with alternative measurements (Fig. 1G and fig. S13); we also found that the range of k_{on} and k_{off} in which the experiment is applicable is far wider than previously recognized (fig. S4).

2) In mammalian proteins, contributions from glycans on the surface of the protein could not be removed from the spectrum by means of relaxation filters used in epitope mapping (*30*) without compromising the sensitivity of the experiment. We addressed this instead by applying baseline subtraction using data obtained from a protein-only sample.

3) The magnetization transfer, and hence the sensitivity of the experiment, will be higher when the excitation frequency of the saturation pulse is close to a maximum in the protein NMR spectrum. If any ligand resonances are outside of the bandwidth of the pulse, and if a "ligand-only" subtraction is applied, the magnetization transfer can be maximized. With this condition, the response for a given protein-ligand system in fact becomes invariant to the excitation frequency used (figs. S9 and S10).

4) In complex molecules, such as sialosides, NMR spectra are crowded and overlapped. To reliably obtain magnetization transfer measurements at all points in the ligand, we developed a peak-picking algorithm based on earlier work (*31*) that can automate the process, returning a list of peak locations and a simulated NMR experiment that can be directly compared to the data. The locations of the peaks are in excellent agreement with the locations for multiplets determined using standard multidimensional approaches used for resonance assignment (see Fig. 1, B and D, Fig. 2D, and Fig. 3A for examples; see also overlaps in all subsequent uSTA analyses and table S7).

5) The uSTA software allowed the combination of intensity from scalar coupled multiplets following a user-input assignment, to provide "per resonance" measures of saturation transfer. These are provided as-is, and also as $\langle 1/r^6 \rangle$ interpolated "binding maps" that represent the interaction on nearby heteroatoms, thereby allowing ready visual inspection of the binding pose of a molecule.

Testing of uSTA in model systems

The uSTA method (Fig. 1, A to C) was tested first in an archetypal, yet challenging, ligand-to-protein interaction (Fig. 1, D to G). Implementation in an automated manner through software governing the uSTA workflow reduced artifacts arising from subjective, manual analyses (fig. S6). The binding of L-tryptophan (Trp) to bovine serum albumin (BSA, Fig. 1D) is a long-standing benchmark (*25*) because of the supposed role of hydrophobicity in the plasticity of this interaction as well as a lack of corresponding fully determined, unambiguous three-dimensional (3D; e.g., crystal) structures. This is also a simpler amino acid–protein interaction system (less-modified protein, small ligand) that classical NMR/STD methods are perceived (*24*, *32*) to have already delineated well.

As for a standard STD experiment, 1D ^{1}H-NMR spectra were determined for both ligand and protein. In addition, mixed spectra containing both protein and excess ligand ($P + L$) were determined with and without excitation irradiation at frequencies corresponding to prominent resonance within the protein but far from any ligand (pulse "on") or where the center of the pulse was moved to avoid ligand and protein (pulse "off," labeled "1D" in the figures). Deconvolved spectra for ligand determined in the presence of protein were matched with high accuracy by uSTA (Fig. 1E). Moreover, uSTA generated highly consistent binding "heatmaps" comprising atom-specific magnetization transfer efficiencies (proton data mapped onto heteroatoms by taking a local $1/r^6$ average to enable visual comparison) that described the pose of ligand bound to protein (Fig. 1F; see also figs. S8 and S11). These were determined over a range of ligand concentrations even as low as 40 μM [Fig. 1, E(iii) and F(iii)] where the ability of uSTA to extract accurate signal proved unprecedented and critical to quantitation of binding (see below). Binding maps were strikingly consistent across concentrations, indicating a single, consistent pose driven by the strongest interaction of protein with the heteroaromatic indole side chain of Trp. This not only proved consistent with x-ray crystal structures of BSA with other hydrophobic ligands, (*33*) it also revealed quantitative subtleties of this interaction at high precision: Protein "grip" is felt more at the distal edge of the indole moiety.

Next, with indications of expanded capability of uSTA in a benchmark system, we moved to first analyses of sugar-protein interactions. Sugar ligand trehalose (Tre) binds only weakly to trehalose repressor protein TreR and so proves challenging in ligand-to-protein interaction analysis (*34*). Nonetheless, the uSTA workflow again successfully and rapidly determined atom-specific transfer efficiencies with high precision and resolution (fig. S3F). Atom-precise subtleties were revealed in this case as well: Hotspots of binding occur around OH-3/OH-4 and graduate to reduced binding around both sugar rings, with only minimal binding of the primary OH-6 hydroxyl (fig. S3F). Once more, this uSTA-mapped $P + L$ interaction proved consistent with prior x-ray crystal structures (*35*).

Direct determination of ligand-protein K_D using uSTA

The precision of signal determination in uSTA critically allowed variation of ligand/protein concentrations even down to low levels (see above), enabling direct determination of binding constants in a manner not possible by classical methods. Following measurement of magnetization transfer between ligand and protein, variation with concentration (Fig. 1E) was quantitatively analyzed using modified Bloch-McConnell equations (*36*) (see Methods). These accounted for intrinsic relaxation, cross-relaxation, and protein-ligand binding (Fig. 1A) to directly provide measurements of equilibrium binding K_D and associated kinetics (k_{ex}). In the Trp/BSA system, this readily revealed $K_D = 38 \pm 15$ μM, $k_{on} = 1.6 (\pm 0.6) \times 10^5$ M^{-1} s^{-1}, and $k_{off} = 6.0 \pm 2.0$ s^{-1} (Fig. 1G), consistent with prior determinations of K_D by other solution-phase methods [$K_D = 30 \pm 9$ μM by isothermal calorimetry (*32*)]. Note that this direct method proved to be possible only because of the ability of the uSTA method to deconvolute a true signal with sufficient precision, even at the lower concentrations used and consequently lower signal (Fig. 1E). Thus, uSTA enabled atom-mapping and quantitation for ligand binding that were improved over previous methods. Critically, these values were fully consistent with all observed NMR data and independently obtained measures of K_D.

uSTA allows interrogation of designed crypticity in influenza HA virus attachment

Having validated the uSTA methodology, we next used it to interrogate sugar binding by viral attachment protein systems that have proved typically intractable to classical methods. The hemagglutinin (HA) trimer of influenza

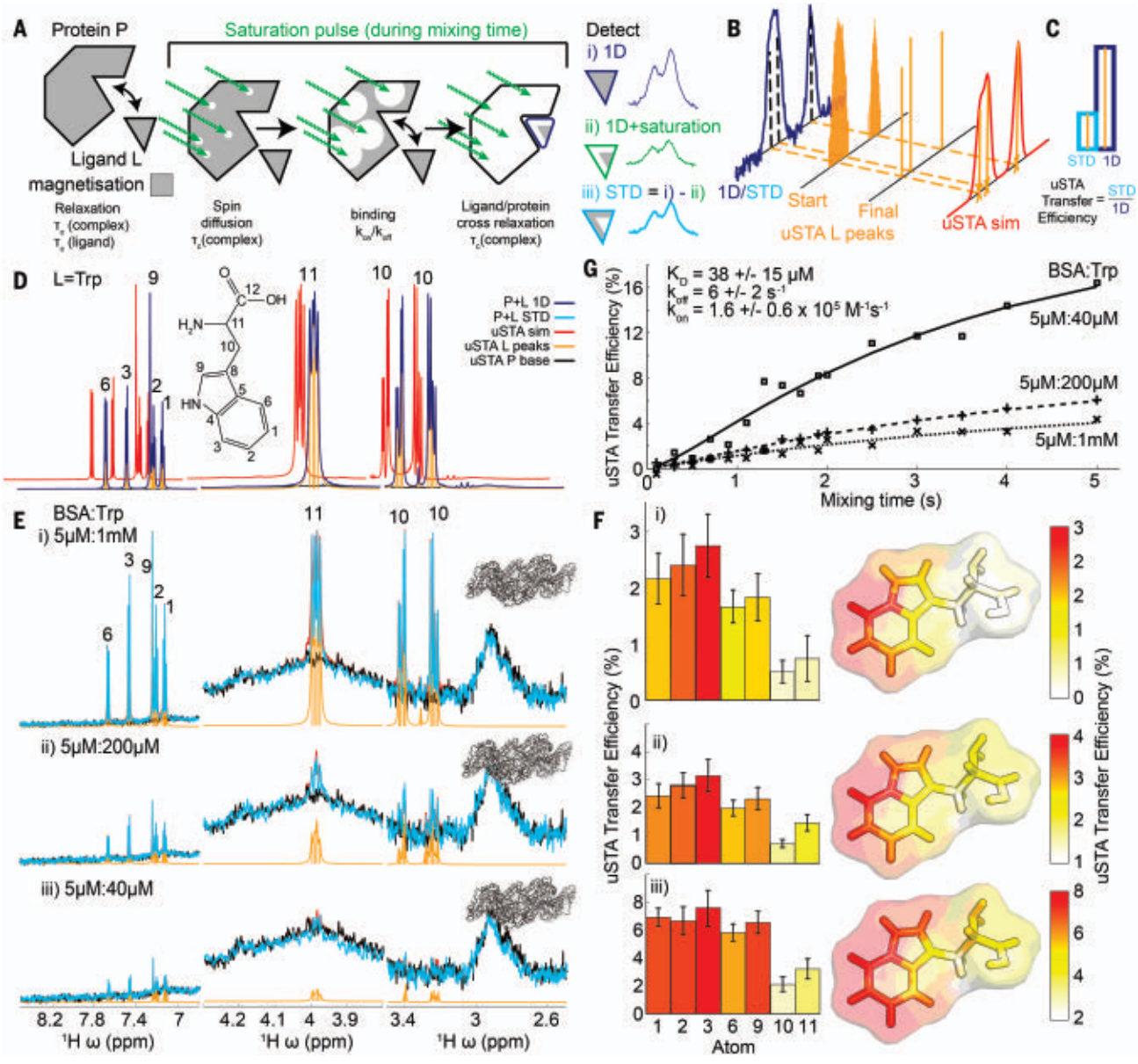


Fig. 1. Development of the uSTA method. (**A** to **C**) Schematic of the process for uSTA that exploits comprehensive numerical analysis of relaxation and ligand binding kinetics (A) using full and automatically quantified signal intensities in NMR spectra (B) and calculates per-resonance transfer efficiencies (C). In (B), signal analysis determined the number of peaks that can give rise to the signal, and returned simulated spectrum by convolving these with peak shape function. The precise peak positions returned are in excellent agreement with the known positions of resonances identified using conventional means. Magnetization transfer NMR experiments compare two 1D NMR spectra, where the second involves a specific saturation pulse that aims to "hit" the protein but "miss" the ligand in its excitation. This is accomplished by acquiring the 1D spectrum with the saturation pulse held off resonance at −35 ppm such that it will not excite protons in either ligand or protein [labeled "1D" in (B) and (C)]. The uSTA method requires these two spectra to be analyzed as described in (B), in pairs, one that contains the raw signal, and the second that is the difference between the two. We define the "transfer efficiency" as the fractional signal that has passed from the ligand to the protein. (**D** to **F**) Application of uSTA to study the interaction between bovine serum albumin (BSA) and L-tryptophan (Trp). In (D), the 1D ^{1}H-NMR spectrum of the mixture at 200 μM Trp and 5 μM BSA (= *P* + *L*, blue) is dominated by ligand, yet the ligand (*L*) and protein (*P*) can still be deconvolved by universal deconvolution, using a reference obtained from a sample containing protein only. This reveals contributions from individual multiplets originating from the ligand (yellow) and the protein-only baseline (black), allowing precise recapitulation of the sum (red). In (E), application of universal deconvolution to STD spectra with varying concentrations of tryptophan allows uSTA using ligand resonances identified in (D). This in turn allows signal intensity in the STD spectrum (*P* + *L* STD, light blue) to be determined with high precision. Although signal-to-noise in the STD increases considerably with increasing ligand concentration, the measured atom-specific transfer efficiencies as determined by uSTA are consistent [(F), left, bar charts; right, transfer efficiency binding "maps"], showing that the primary contact between protein and ligand occurs on the distal edge (C-1, 2, 3, 4; N-7 and C-9 using the numbering shown) of the indole aromatic ring. Application of the same uSTA workflow also allowed precise determination of even weakly binding sugar ligand trehalose (Glc-α1, 1α-Glc) to *E. coli* trehalose repressor TreR. Again, uSTA allows determination of transfer efficiencies with atom-specific precision (see fig. S3F). (**G**) Quantitative analysis of the STD build-up curves using a modified set of Bloch-McConnell equations that account for binding and cross-relaxation allows us to determine thermodynamic and kinetic parameters that describe the BSA-Trp interaction, K_D, k_{on}, and k_{off}. The values obtained are indicated and are in excellent accord with those obtained by other methods (*25*, *32*). Errors come from a bootstrapping procedure (see Methods).

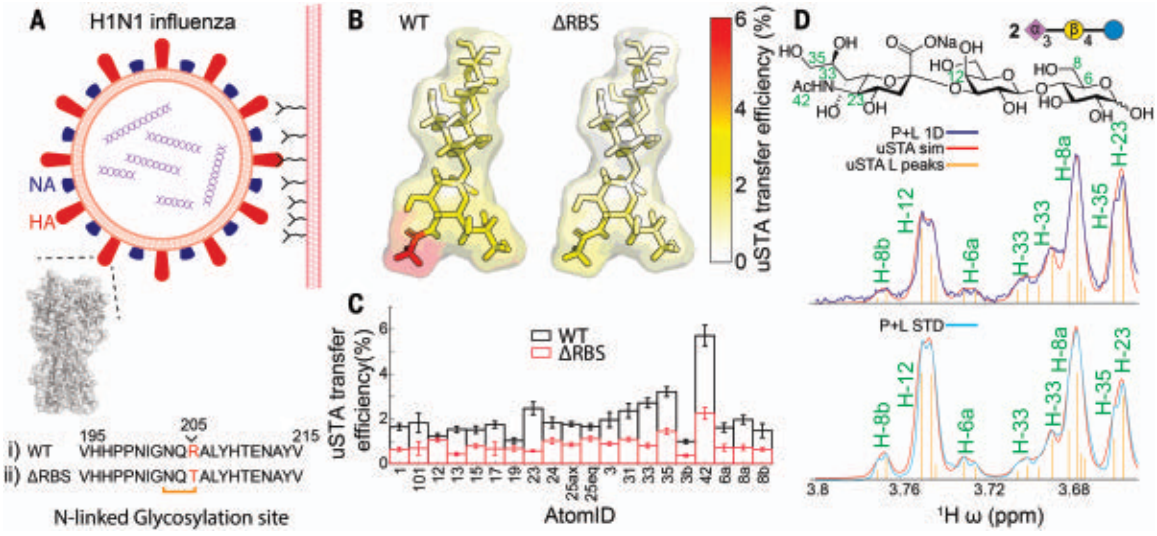


Fig. 2. uSTA allows mapping of a designed cryptic sugar-binding site in H1N1 influenza hemagglutinin (HA). (**A**) HA presents on the surface of the viral membrane and has been shown to bind with sialic acid surface glycans to mediate host cell entry. A designed (*38*) ΔRBS variant is generated by the creation of an N-linked glycosylation site via the creation of needed sequon NQT from wild-type NQR by the R205T point mutation in HA adjacent to the sialic acid–binding site. In this way, disruption of sialic acid binding through designed "blocking" in HA-ΔRBS is intended to ablate the binding of HA wild-type to sialosides in this synthetic variant. (**B** and **C**) Notably, wild-type and ΔRBS variants of H1N1 HA in fact show a remarkably similar overall binding pattern (B) for the 2,3-sialo-trisaccharide **2** (focused through engagement with the sialoside) but a significant intensity moderation (C) for the ΔRBS variant, indicative of a partial (but not complete) loss of binding consistent with design (*38*). (**D**) Raw spectral data demonstrate that atom-specific differences in intensity in the 1D versus the difference spectra can be discerned using uSTA. Note that atom numbering shown and used here is generated automatically by uSTA.

A virus is known to be essential for its exploitation of sialoside binding (*37*); H1N1 has emerged as one of the most threatening variants in recent years. We took the H1 HA in both native form and a modified form, containing a non-natural sequon specific for N-glycosylation that was previously designed (*38*) to block intermolecular (in trans) sugar binding. This designed blocking in a so-called HA-ΔRBS variant (*38*) also notably creates an additional glycan beyond the existing, potentially confounding, glycosylation background. It therefore provided another test of uSTA's ability to delineate relevant sialoside ligand interactions in another important pathogen protein (Fig. 2A). Despite this intended blocking, the precision of uSTA was such that residual in trans binding of sialo-trisaccharide **2** could still be detected in HA-ΔRBS, albeit at a lower, modulated level [as expected by design (*38*)]. Although H1 HA is known to bind both sialo-trisaccharides **2** and **3**, **2** is the less preferred (2,6- over 2,3-linked) ligand (*39*), and yet its binding could still be mapped (here, to a mode mediated primarily by the sialoside moiety). In this way, the sensitivity of uSTA to detect even lower sialoside binding to relevant proteins was confirmed.

uSTA reveals natural, cryptic sialoside binding by SARS-CoV-2 spike

We next probed putative, naturally cryptic sialoside binding sites in SARS-CoV-2. Our analysis of the 1D protein ^{1}H-NMR spectrum of the purified prefusion-stabilized ectodomain construct (*10*) of intact trimeric SARS-CoV-2 spike attachment protein (fig. S7) revealed extensive protein glycosylation with sufficient mobility to generate a strong ^{1}H-NMR resonance in the region 3.4 to 4.0 ppm (Fig. 3A). Although lacking detail, these resonances displayed chemical shifts consistent with the described mixed patterns of oligomannose, hybrid, and complex N-glycosylation found on SARS-CoV-2 spike after expression in human cells (*40*). As such, these mobile glycans on SARS-CoV-2 spike contain sialoside glycan residues that not only confound analyses by classical NMR methods but are also potential competing, "internal" (in cis) ligands for any putative attachment (in trans) interactions, as well as possible direct ligands for in trans interactions in their own right (*41*). Therefore, their presence in the protein NMR analysis presented clear confounding issues for typical classical STD analyses. As such, SARS-CoV-2 spike represented a stringent and important test of the uSTA method.

We used uSTA to evaluate a representative panel of both natural and site-specifically modified unnatural sialosides as possible ligands of spike (Fig. 3 and figs. S3 and S12). Use of classical methods provided an ambiguous assessment (fig. S8), but use of uSTA immediately revealed binding and nonbinding sugar ligands (Fig. 3, figs. S8, S10, and S11, and table S7). Initially, the simplest sialoside, *N*-acetyl-neuraminic acid (**1**), was tested as a mixture of its mutarotating anomers (**1α** ⇔ **1β**). When analyzed using uSTA, these revealed [Fig. 3, E and F(i, ii), table S7, and fig. S11] clear "end-on" interactions by **1α** as a ligand [Fig. 3F(i)], mediated primarily by the acetamide NHAc-5, but no reliably measurable interactions by **1β** [Fig. 3F(ii)]. This detection of selective α-anomer interaction, despite the much greater dominance of the β-anomer in solution, provided yet another demonstration of the power of the uSTA method, here operating in the background of dominant alternative sugar (Fig. 3E and fig. S11). The α selectivity correlates with the near-exclusive occurrence of sialosides on host cell surfaces as their α- but not β-linked conjugates (see also below).

Having confirmed simple, selective monosaccharide α-sialoside binding, we explored extended α-sialoside oligosaccharide ligands (compounds **2** and **3**; Fig. 3, C and D) that would give further insight into the binding of natural endogenous human cell-surface sugars as well as unnatural variants (compounds **4** to **6**; Fig. 3, F and G) that could potentially interrupt such binding. Sialosides are often found appended to galactosyl (Gal/GalNAc) residues in either α2,3-linked (**2**) or α2,6-linked form (**3**). Both were tested (Fig. 3, C and D) and exhibited "end-on" binding consistent with that seen for *N*-acetyl-neuraminic acid (**1**) alone, but with more extended binding surfaces (Fig. 3D), qualitatively suggesting a stronger binding affinity (see below for quantitative analysis). Common features of all sialoside binding modes were observed: The NHAc-5

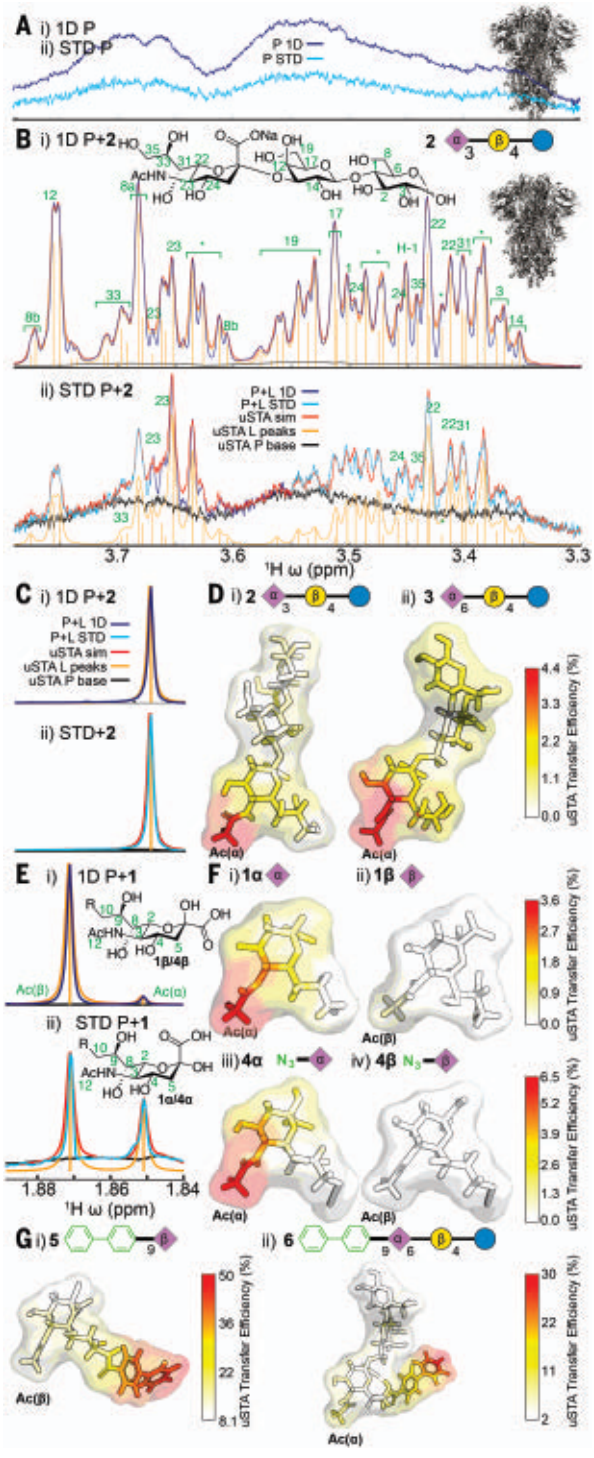


Fig. 3. uSTA reveals interaction of sialosides with SARS-CoV-2 spike protein. A panel of natural, unnatural, and hybrid variant sialoside sugars **1–6** (see fig. S12) was used to probe interaction between sialosides and spike. (**A**) The 1D ^{1}H-NMR of SARS-CoV-2 spike protein shows considerable signal in the glycan-associated region despite protein size, indicative of mobile internal glycans in spike protein. This effectively masks traditional analyses, as without careful subtraction of the protein's contributions to the spectrum (fig. S8), the ligand cannot be effectively studied. (**B**) Application of the uSTA workflow (fig. S6) to SARS-CoV-2 spike protein (shown in detail for **2**). The uSTA process of ligand peak assignment and deconvolution → *P* + *L* peak assignment and deconvolution → application to *P* + *L* STD yields precise atom-specific transfer efficiencies (fig. S6). Note how in (ii) individual multiplet components, have been assigned (yellow); the back-calculated deconvolved spectrum (red) is an extremely close match for the raw data (purple). The spectrum is a complex superposition of the ligand spectrum (and protein only yet uSTA again accurately deconvolves the spectrum, revealing the contribution of protein-only (black) and the ligand peaks (yellow). Using these data, uSTA analysis of the STD spectrum pinpoints ligand peaks and signal intensities. Spectral atom numbering shown and used here is generated automatically by uSTA; all other numbering in sugars follows carbohydrate nomenclature convention. (**C** and **D**) Application of the uSTA workflow (fig. S6) reveals atom-specific binding modes to spike protein for both natural (e.g., sialoside) trisaccharides Siaα2,3Galβ1,4Glc (**2**) and Siaα2,6Galβ1,4Glc (**3**). Comparison of the uSTA method focused on the NHAc methyl resonance shows excellent agreement (C). The uSTA method allowed determination of binding surfaces for both trisaccharides **2** [D(i)] and **3** [D(ii)]. (**E** and **F**) STD spectrum (E) and mapped atom-specific transfer efficiencies (F) for sialic acid (**1**) and 9-N_3 azido variant **4**. Both interconverting α and β anomeric forms could be readily identified. Despite the dominance of the β form [94 %, E(i)], application of the uSTA method following assignment of resonances from the two forms allowed determination of binding surfaces simultaneously [E(ii), F(i, ii)]. Spike shows strong binding preference for the α anomers [F(i, iii) versus F(ii, iv)] despite this strong population difference. Binding surfaces were also highly similar to those of extended trisaccharides **2** and **3**. (**G**) Using these intensities, atom-specific transfer efficiencies can be determined with high precision, shown here for hybrid sialoside **5**. The details of both the unnatural BPC moiety and the natural sialic acid moiety can be mapped; although the unnatural aromatic BPC dominates interaction, uSTA nonetheless delineates the subtleties of the associated contributions from the natural sugar moiety in this ligand (see also figs. S5 and S6).

acetamide of the terminal sialic acid (Sia) is a binding hotspot in **1**, **2**, and **3** that drives the "end-on" binding. Differences were also observed: The α2,6-trisaccharide (**3**) displayed a more extended binding face yet with less intense binding hotspots (Fig. 3D) engaging additionally the side-chain glycerol moiety (C7–C9) of the terminal Sia acid as well as the OH-4 C4 hydroxyl of the Gal residue. The interaction with α2,3-trisaccharide (**2**) was tighter and more specific to NHAc-5 of the Sia.

These interactions of the glycerol C7–C9 side chain detected by uSTA were probed further through construction (fig. S12) of unnatural modified variants (**4** to **6**; Fig. 3, F and G). Replacement of the OH-9 hydroxyl group of sialic acid with azide N_3-9 in **4** [Fig. 3F(iii) and table S7] was well tolerated, but larger changes (replacement with aromatic group biphenylcarboxamide BPC-9) in **5** and **6** led to an apparently abrupt shift in binding mode that was instead dominated by the unnatural hydrophobic aromatic modification (Fig. 3G and Fig. 4C). As for native sugar **1**, azide-modified sugar **4** also interacted with spike in a stereochemically specific manner with only the α-anomer displaying interaction [Fig. 3F(iii)], despite dominance of the β-anomer in solution [Fig. 3F(iv)]. uSTA allowed precise dissection of interaction contributions in these unusual hybrid (natural-unnatural) sugar ligands that could not have been determined using classical methods (see text S5).

Using variable concentrations of the most potent natural ligand α2,3-trisaccharide **2** [6 μM spike, **2** at 60 μM, 200 μM, 1 mM, and 2 mM excitation at 5.3 ppm] and variable concentrations of spike protein, we used the uSTA method to directly determine solution-phase affinities (Fig. 4A): K_D = 32 ± 12 μM, k_{on} = 6300 ± 2300 M^{-1} s^{-1}, and k_{off} = 0.20 ± 0.08 s^{-1}. We also probed binding in a different mode by measuring the affinity of spike to **2** when displayed on a modified surface (fig. S13) using surface plasmon resonance (SPR) analysis. The latter generated a corresponding K_D = 23.7 ± 3.6 μM (k_{on} = 1004 ± 290 M^{-1} s^{-1}). Such similar values for sialoside ligand in solution (by uSTA) or when displayed at a solid-solution interface (by SPR) suggested no substantial avidity gain from display of multiple sugars on a surface.

Structural insights from uSTA delineate binding to SARS-CoV-2 spike

uSTA analyses consistently identified binding hotspots in sugars **1** to **4** providing the highest transfer efficiencies in an atom-specific manner, particular the "end" NHAc-5 acetamide methyl group of the tip sialic acid residue in all. A combination of uSTA with so-called high ambiguity–driven docking (HADDOCK) methods (*42*, *43*) was then used to probe likely regions in SARS-CoV-2 spike for this "end-on" binding mode via uSTA data-driven atomistic

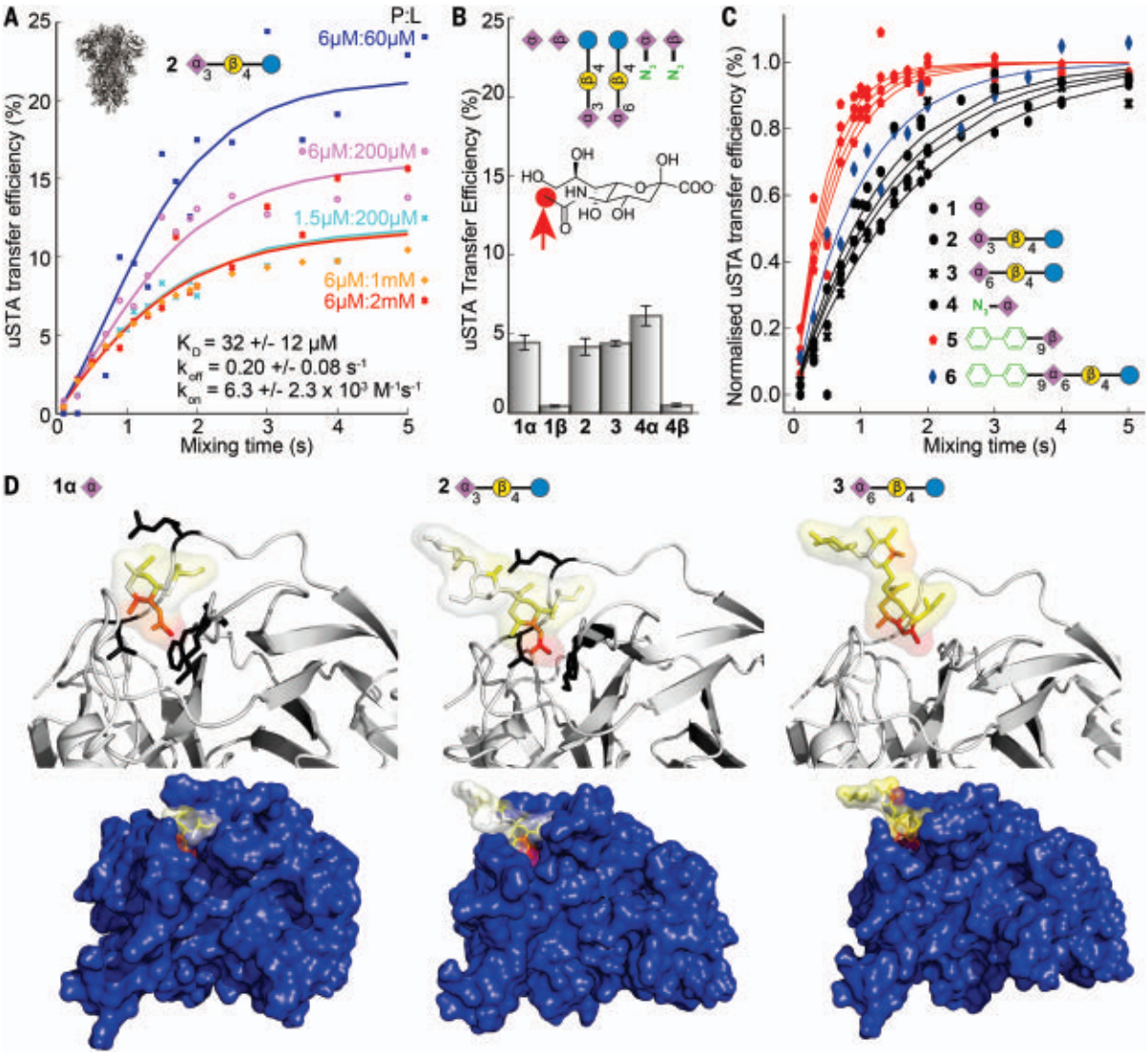


Fig. 4. Quantitative uSTA analyses allow comparison and predictive restraints for protein ligand-binding prediction. (**A**) Quantitative analysis of the STD build-up curves using a modified set of Bloch-McConnell equations that account for binding and cross-relaxation allow us to determine thermodynamic and kinetic parameters that describe the SARS-CoV-2•**2** interaction, K_D, k_{on}, and k_{off}. The values obtained are indicated. Errors come from a bootstrapping procedure (see Methods). The lower ligand concentrations yield data that are of lower sensitivity than higher concentrations. The transfer efficiencies will be higher in this case, as more molecules are effectively involved in the binding. Thus, data at lower concentrations will in general have more scatter and higher transfer efficiencies. These data points are desirable for the analysis, as it is here where we expected the greatest variation of transfer efficiency with ligand concentration. The analysis is applied globally and so the uncertainties in the final fitted parameters from the bootstrapping analysis (see Methods) provide a direct and confident measure of the goodness of fit. (**B**) Normalized uSTA transfer efficiencies of the NAc-5 methyl protons can be determined for each ligand studied here. This allowed relative contributions to "end on" binding to be assessed via uSTA in a "mode-specific" manner. This confirmed strong α- over β-sialoside selectivity. Errors were determined through a bootstrapping procedure where mixing times were sampled with replacement, allowing for the construction of histograms of values in the various parameters that robustly reflect their fitting errors. (**C**) Normalized build-up curves for the most intense resonances allowed two distinct modes of binding to be identified in natural (**1–4**) and hybrid (**5**, **6**) sugars. Data are shown at constant protein and ligand concentrations. With the BPC moiety present, the build-up of magnetization occurs significantly faster than when not; various such hybrid ligands give highly similar curves. By contrast, natural ligands have a much slower build-up of magnetization. This, together with the absolute transfer efficiencies being very different, and the overall pattern on the interaction map combine to reveal that the ligands are most likely binding via two different modes and possibly locations on the protein. (**D**) Coupling uSTA with an integrative modeling approach such as HADDOCK (*42*, *43*) allowed generation and, by quantitative scoring against the experimental uSTA data, selection of models that provide atomistic insights into the binding of sugars to the SARS-CoV-2 spike protein, as shown here by superposition of uSTA binding "map" onto modeled poses. uSTA mapping the interaction between SARS-CoV-2 spike [based on RCSB 7c2l (*19*)] with ligands **1a**, **2**, and **3** identifies the NHAc-5 methyl group of the tip sialic acid residue making the strongest interaction with the protein. By filtering HADDOCK models against this information, we obtain structural models that describe the interaction between ligand and protein (fig. S14). Most strikingly, we see the same pattern of interactions between protein and sialic acid moiety in each case, where the NAc methyl pocket is described by a pocket in the spike NTD. Although sequence and structural homology are low (fig. S1), MERS spike protein possesses a corresponding NHAc-binding pocket characterized by an aromatic (Phe^{39})–hydrogen-bonding (Asp^{36})–hydrophobic (Ile^{132}) triad (*5*).

models. In each case, a cluster of likely poses emerged (Fig. 4D) for **1α**, **2**, and **3** (see fig. S14 for details) consistent with "end-on" binding where the acetamide NHAc-5 methyl group of the sialic acid moiety was held by the unusual β sheet–rich region of the NTD of SARS-CoV-2 spike. Under the restraints of uSTA and homology, a glycan-binding pocket was delineated by a triad of residues (Phe^{79}, Thr^{259}, and Leu^{18}) mediating aromatic, carbonyl hydrogen–bonding, and hydrophobic interactions, respectively. However, the sequence and structural homology to prior (i.e., MERS) coronavirus spike proteins in this predicted region was low; the MERS spike protein uses a corresponding NHAc-binding pocket characterized by an aromatic (Phe^{39}), hydrogen-bonding (Asp^{36}), and hydrophobic (Ile^{132}) triad to bind the modified sugar 9-*O*-acetyl-sialic acid (*5*).

SARS-CoV-2 glycan attachment mechanisms have to date only identified a role for spike RBD in binding rather than NTD (*15*, *17*). We used uSTA to compare the relative potency of the sialoside binding identified here to previously identified (*17*) heparin binding motifs (Fig. 5, A and B). Heparin sugars **7** and **8** of similar size to natural sialosides **3** and **4** were selected so as to allow a near ligand-for-ligand comparison based on similar potential binding surface areas. **7** and **8** also differed from each other only at a single glycan residue (residue 2) site to allow possible dissection of subtle contributions to binding. Unlike the "end-on" binding seen for sialosides **3** and **4**, uSTA revealed an extended, nonlocalized binding interface for **7** and **8** consistent instead with "side-on" binding [Fig. 5, A(ii) and B(ii)].

Next, we examined the possible evolution of sialoside binding over lineages of SARS-CoV-2 (*44*). Four notable variants of concern—alpha/B1.1.7, beta/B1.351, delta/B.1.617.2, and omicron/B.1.1.529—emerged in later phases of the pandemic. When these corresponding spike protein variants were probed by uSTA, all displayed ablated binding toward sialoside **2** as compared to first-phase B-origin-lineage spike (Fig. 5D and fig. S15).

Finally, to explore the possible role of sialoside binding in relation to ACE2 binding, we also used uSTA to probe the effects upon binding of the addition of a known, potent neutralizing antibody of ACE2-spike binding, C5 (Fig. 5, D and E, and fig. S16) (*45*, *46*). Assessment of binding to sialoside **2** in the presence and absence of antibody at a concentration sufficient to saturate the RBD led to only slight reduction in binding. Uniformly modulated atomic transfer efficiencies and near-identical binding maps (Fig. 5, E and F) were consistent with a maintained sialoside-binding pocket with undisrupted topology and mode of binding.

Together, these findings allow us to conclude that the sialoside binding observed with uSTA involves a previously unidentified "end-on" mechanism/mode that operates in addition to and potentially cooperatively with

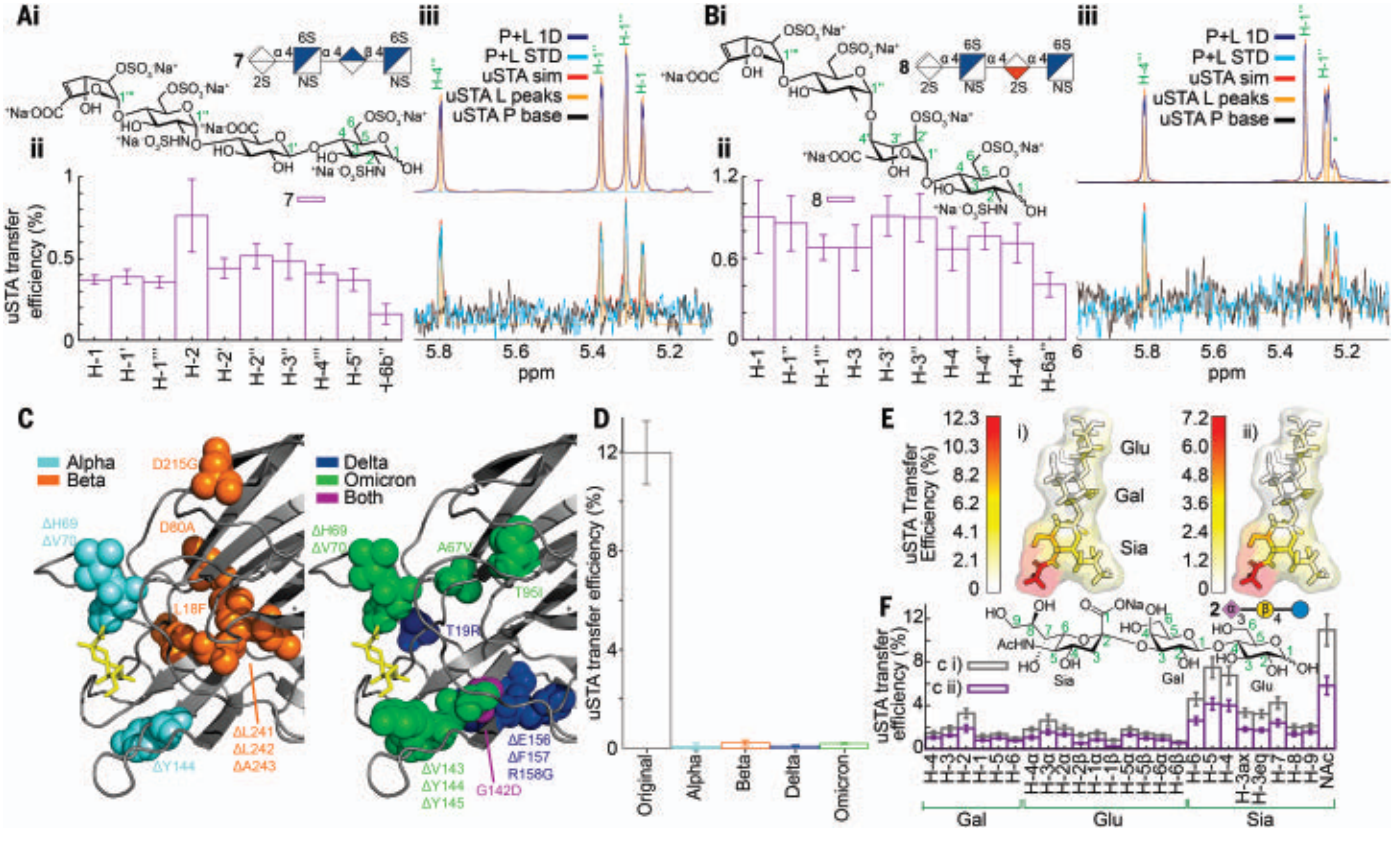


Fig. 5. Comparison of SARS-CoV-2 glycan attachment mechanisms and variant evolution via uSTA suggests binding away from the RBD that is lost. (**A** and **B**) Two heparin tetrasaccharides (**7**, **8**) are shown by uSTA [A(iii), B(iii)] to bind B-origin-lineage SARS-CoV-2 spike protein ("original" spike) in a "side-on" mode [A(ii), B(ii)]. Atom specific binding is shown in A(ii) and B(ii). Assignments shown in green use conventional glycan numbering. (**C**) Substantial numbers of mutations arise in the NTD region identified by uSTA in the B.1.1.7/alpha (cyan), B.1.351/beta (orange), B.1.617.2/delta (dark blue) and B.1.1.529/omicron (green) lineage variants of SARS-CoV-2 spike. (**D**) Ablated binding of the α-2,3-sialo-trisaccharide **2**, as measured by the transfer efficiency of the NHAc protons, is identified by uSTA in the lineage variants of SARS-CoV-2 spike [see (C) for colors]. This proves consistent with mutations that appear in the sialoside-binding site of NTD identified in this study [see (C) and Fig. 6]. (**E** and **F**) uSTA of sialoside **2** with B-origin-lineage SARS-CoV-2 spike in the presence of the potent RBD-neutralizing nanobody C5 [spike E(i), nanobody-plus-spike E(ii)] shows essentially similar binding patterns with uniformly modulated atomic transfer efficiencies (F).

ACE2 binding in SARS-CoV-2. The primary sialoside glycan-binding site SARS-CoV-2 spike is distinct from that of heparin ("end-on" versus "side-on"), not in the RBD (not neutralized by RBD-binding antibody), and found instead in an unusual NTD region that has become altered in emergent variants (loss of binding in alpha and beta variants of concern).

Cryo-EM pinpoints the sialoside-binding site in B-origin-lineage spike

Structural analysis of the possible binding of sialosides has been hampered to date by the moderate resolution, typically less than 3 Å, of most SARS-CoV-2 spike structures. Currently deposited cryo-EM–derived coulombic maps show that the NTD of spike is often the least well-resolved region. In our initial attempts with native protein, large stretches of amino acids within the NTD were not experimentally located (*45*); the most disordered regions occur in the NTD regions that contribute to the surface of the spike. A stabilized closed mutant form (*47*) of the spike was examined and gave improvements, but the coulombic map was still too weak and noisy to permit tracing of many of the loops in the NTD. However, with a reported fatty acid–bound form of the spike (*48*), which has shown prior improved definition of the NTD, we were able to collect a 2.3 Å dataset in the presence of the α2,3-sialo-trisaccharide **9**. The map was clear for almost the entire structure including the previously identified linoleic acid (fig. S17D); only 13 N-terminal residues and two loops (residues 618 to 632 and 676 to 689) were not located. Although the density is weaker at the outer surface of the NTD than at the core of the structure (fig. S17B), the map was of sufficient quality to model N-glycosylation at site Asn^{149}, which is in a flexible region, and the fucosylation state of N-linked glycans at Asn^{165} (Fig. 6A).

We observed density in a pocket at the surface of the NTD lined by residues His^{69}, Tyr^{145}, Trp^{152}, Gln^{183}, Leu^{249}, and Thr^{259} (Fig. 2B and fig. S17). This density is absent in other spike structures of higher than 2.7 Å resolution (PDB IDs 7jji, 7a4n, 7dwy, 6x29, 6zge, 6xlu, 7n8h, 6zb5, and 7lxy), even those (such as PDB IDs 7jji, 6zb5, and 6zge) that have a well-ordered NTD. The density when contoured at 2.6σ is fitted by an α-sialoside consistent with the terminal residue of **9**, with the distinctive glycerol and N-acetyl groups clear. To further strengthen our confidence in the identification of the sialic acid, we determined a native (unsoaked) structure to 2.4 Å using the same batch of protein (fig. S17D). This structure showed no density in the sialic acid binding site supporting our assignment. The sialoside was therefore included in the refinement and the thermal factors (108 $Å^2$) were comparable to those for the adjacent protein residues (95 to 108 $Å^2$). In this position, the glycerol moiety makes a hydrogen bond with the side chain of Tyr^{145}, the N-acetyl group with Ser^{247}, and the carboxylate with Gln^{183}, and there are hydrophobic interactions with Trp^{152}. Lowering the map threshold to 1.6σ would be consistent with a second pyranoside (e.g., galactoside) residue (fig. S17C). At this contour level, the map clearly covers the axially configured carboxylate of the sialic acid (fig. S17). The middle galactoside residue of **9** positioned in this density would make contacts with Arg^{248} and Leu^{249}.

Structural superposition of the NTD with that of the MERS spike (RCSB 6NZK) shows that although the sialic acid–binding pockets of both are on the outer surface, these pockets are 12 Å apart (as judged by the C2 atom of respective sialic acids; Fig. 6D). In MERS spike, sialic acid is bound at the edge of the central β sheet, whereas in SARS-CoV-2 the sugar is bound at the center of the sheet; thus, the pockets use different elements of secondary structure. Because of distinct changes in the structure of the loops connecting the strands, the sialic acid pocket from one protein is not present in the other protein. Several regions of additional density were not fitted by the model (fig. S18).

Disclosure of sialoside trisaccharide as a ligand for B-origin-lineage SARS-CoV-2 correlates with clinical genetic variation in early-phase pandemic

A distinctive mode of sialoside binding by spike confirms a potential attachment point for SARS-CoV-2 found commonly on cell surfaces (sialosides are attached both as glycolipid and glycoprotein glycoconjugates), thus raising the question of whether glycosylation function in humans affects infection by SARS-CoV-2 and hence the presentation and pathology of COVID-19 disease. Analysis of whole-exome sequencing data of an early 2020 cohort of 533 COVID-19–positive patients (see table S1) identified two glycan-associated genes within the top five that were most influential upon disease severity. Specifically, recursive feature elimination applied to a LASSO (least absolute shrinkage and selection operator)–based (*49*) logistic regression model identified *LGALS3BP* (fourth of >18,000 analyzed genes) and *B3GNT8* (fifth of >18,000) (Fig. 7A and fig. S19). Variants in these two genes were beneficially associated with less severe disease outcome (Fig. 7, B and C; see also tables S2 to S6 for specific *B3GNT8* and *LGALS3BP* genetic variants, *B3GNT8* χ^2 five categories, *B3GNT8* χ^2 2×2, *LGALS3BP* χ^2 five categories, *LGALS3BP* χ^2 2×2, respectively).

LGALS3BP encodes for a secreted protein, galectin-3–binding protein (Gal-3-BP, also known as Mac-2-BP), that is a partner and blocker of a specific member (Gal-3) of the galectin class of carbohydrate-binding proteins (*50*). Galectins are soluble and are typically secreted and implicated in a wide range of cellular functions (*51*). Notably, Gal-3 binds the so-called poly-*N*-acetyl-lactosamine [polyLacNAc or $(Gal\text{-}GlcNAc)_n$] chain-extension variants found in tetraantennary N-linked glycoproteins (Fig. 7D), including those displaying sialyl-Gal-GlcNAc sialoside motifs (*52*, *53*). Variants in *LGALS3BP* were present in 9 of 114 a/pauci-symptomatic subjects or mildly affected patients (~8%) compared to 8 of the remaining 419 patients who required more intensive care: oxygen support, CPAP/BiPAP, or intubation (<2%); none of the 69 most seriously affected patients (intubated) carried variants of *LGALS3BP* (Fig. 7B). Intact *LGALS3BP* gene product Gal-3-BP therefore appears correlated with more severe COVID-19 outcome. The other implicated gene, *B3GNT8*, encodes a protein glycosyltransferase, β-1,3-*N*-acetyl-glucosaminyltransferase-8 (GlcNAcT8 or β3GnT8), that is responsible for the creation of the anchor point of poly-*N*-acetyl-lactosamine (polyLacNAc) in such tetra-antennary N-linked glycoproteins (Fig. 7D) (*54*). Again, rare variants in *B3GNT8* were present in 11 of 114 of a/pauci-symptomatic subjects or mildly affected patients (~10%) compared to 10 of the remaining 419 patients who required more intensive care (~2%) (Fig. 7C).

Discussion

Experimentally, there are still few, if any, orthogonal approaches to the useful surface display methods (e.g., "glycoarrays") currently used for readily surveying ligands that might be exploited by pathogens. Following validation in model ligand-protein systems, uSTA provided a ready method for identifying sugar ligands bound by pathogens, as well as their binding parameters and poses, even in posttranslationally modified (e.g., glycosylated) protein systems.

In an influenza virus HA protein variant designed to abolish binding through competition by an added glycan site on HA (*55*), uSTA was nonetheless able to unambiguously reveal and "map" residual sialoside binding despite the presence of an added protein-linked glycan as "internal blocker." This is a protein type that has been well-studied in array formats (*39*); we showed here that, even with a "blocked" HA (in a glycosylated state) and a non-preferred 2,3-sialoside ligand, binding could still be mapped by uSTA.

In the spike trimer of the B-origin lineage of SARS-CoV-2, despite the presence of mobile, protein-linked glycans, uSTA clearly revealed sialoside binding and, through mapping, revealed that this binding is more potent when the sialosyl moiety terminates galactosyl oligosaccharides. This pose is in agreement with our cryo-EM structures, which show that the NHAc-5 N-acetyl group at the sugar's tip is buried, a mode of binding we refer to as "end-on." Prior modeling was partly misled by use of lower-resolution structures of spike, because the NTD is highly disordered in these

initial structures. uSTA and cryo-EM also identified a second, differing mode of binding in SARS-CoV-2 spike only by hybrid, aromatic sugars (e.g., **5**, **6**, **9**) driven by aromatic engagement, but the physiological relevance of this binding pocket is currently unclear. We cannot exclude the presence of another sialic acid–binding site (*15*), but there is no structural support for it.

The spike sialoside-binding site in the NTD is also coincident or near to numerous mutational and deletion hotspots in, for example, alpha and omicron (His^{69}, Val^{70}, Tyr^{145}) and beta (β strand Leu^{241}-Leu^{242}-Ala^{243}-Leu^{244}) variants of concern (Fig. 5C). Changes remove either key interacting residues (alpha/omicron variant: residues 69, 70, 145) or perturb structurally important residues (beta variant: β strand) that form the pocket. The "end-on" binding we observed is quite different to the "side-on" binding observed for heparin (*17*). Heparins are often bound in a non–sequence-specific, charge-mediated manner, consistent with such a "side-on" mode. The location of three binding sites in the trimer, essentially at the extreme edges of the spike (Fig. 6C), also imposes substantial geometric constraints for avidity enhancement through multivalency (*56*, *57*).

We also find a clear link between our data and genetic analyses of patients that correlate with the severity of their disease. This association suggests potential roles in infection and disease progression for cell-surface glycans and the two glycan-associated genes that we have identified. Despite their independent identification here, both gene products interact around a common glycan motif: the polyLacNAc chain-extension variants found in tetraantennary N-linked glycoproteins. Consistent with the sialoside ligands found here, these glycoproteins contain Sia-Gal-GlcNAc motifs within N-linked polyLacNAc chains. These motifs have recently been identified in the deeper human lung (*58*).

These data lead us to suggest that B-origin-lineage SARS-CoV-2 virus may have exploited glycan-mediated attachment to host cells (Fig. 7D) using N-linked polyLacNAc chains as a foothold. Reduction of Gal-3-BP function would allow its target, the lectin Gal-3, to bind more effectively to N-linked polyLacNAc chains, thereby competing with SARS-CoV-2 virus. Similarly, loss of β3GnT8 function would ablate the production of foothold N-linked polyLacNAc chains, directly denying the virus a foothold. We cannot exclude other possible mechanisms including, for example, the role of N-linked polyLacNAc chains in T cell regulation (*59*) or glycolipid ligands (*15*). This analysis of the influence of genetic variation upon susceptibility to virus was confined to "first wave," early-pandemic patients infected with B-origin-lineage SARS-CoV-2. Our discovery here also that in B-lineage virus such

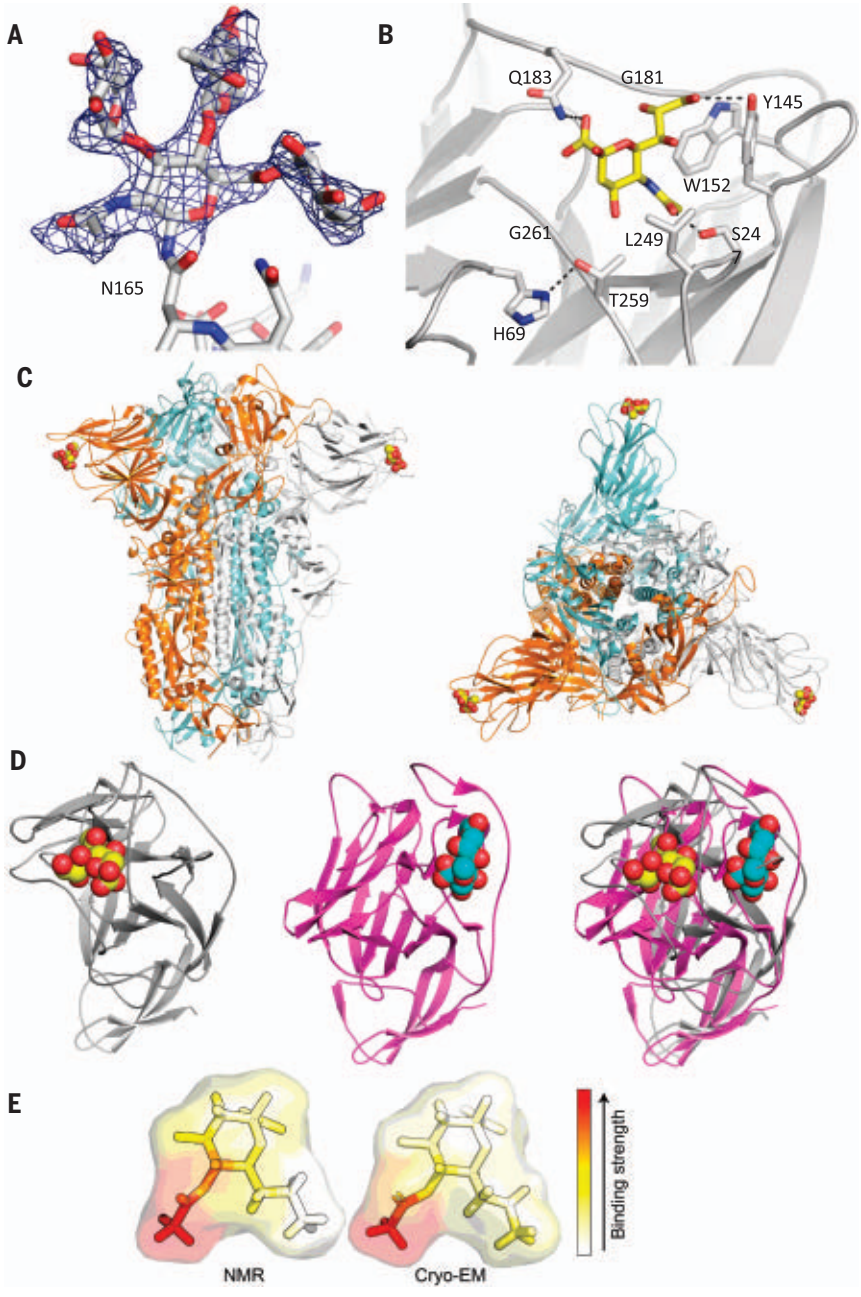


Fig. 6. Cryo-EM analysis of sialoside binding in B-origin-lineage SARS-CoV-2 spike. (**A**) In the presence of sialoside **9** (fig. S18), a well-ordered structure with resolution (2.3 Å) sufficient to identify even glycosylation states of N-linked glycans within spike was obtained. Here they reveal a paucimannosidic base bis-fucosylated chitobiose core structure GlcNAcβ1,4(Fucα1,3-)(Fucα1,6-)GlcNAc–Asn. (**B**) The sialoside-binding site in the NTD is bounded by His^{69}, Tyr^{145}, Trp^{152}, Ser^{247}, and Gln^{183}. Gln^{183} controls stereochemical recognition of α-sialosides by engaging COOH-1. Tyr^{145} engages the C7-C9 glycerol side chain of sialoside. Ser^{247} engages the NHAc-5. Notably, Tyr^{145} and His^{69} are deleted in the alpha variant of SARS-CoV-2 that loses its ability to bind sialoside (see Fig. 5). See also fig. S17 for coulombic maps. (**C**) Spike's sialoside binding site is found in a distinct region of the NTD (left, from side, right from above) that is at the "edge" of the spike (see text). (**D**) Superposition (right) of SARS-CoV-2-NTD (left, gray) with MERS-NTD (middle, magenta) shows distinct sites 12 Å apart. (**E**) A comparison of the ligand-binding mode measured by uSTA NMR and a map calculated from the cryo-EM structure. For each hydrogen environment in the resolved sialic acid, an array of distances to all protein hydrogens, r, was calculated. The interaction of each ligand environment with the spike protein was defined as the summation of all $1/r^6$ environment-protein distances. Values were interpolated in a $1/r^6$ manner onto heteroatoms following the same procedure according to the described NMR methods, and the color bar was scaled from zero to the maximum value.

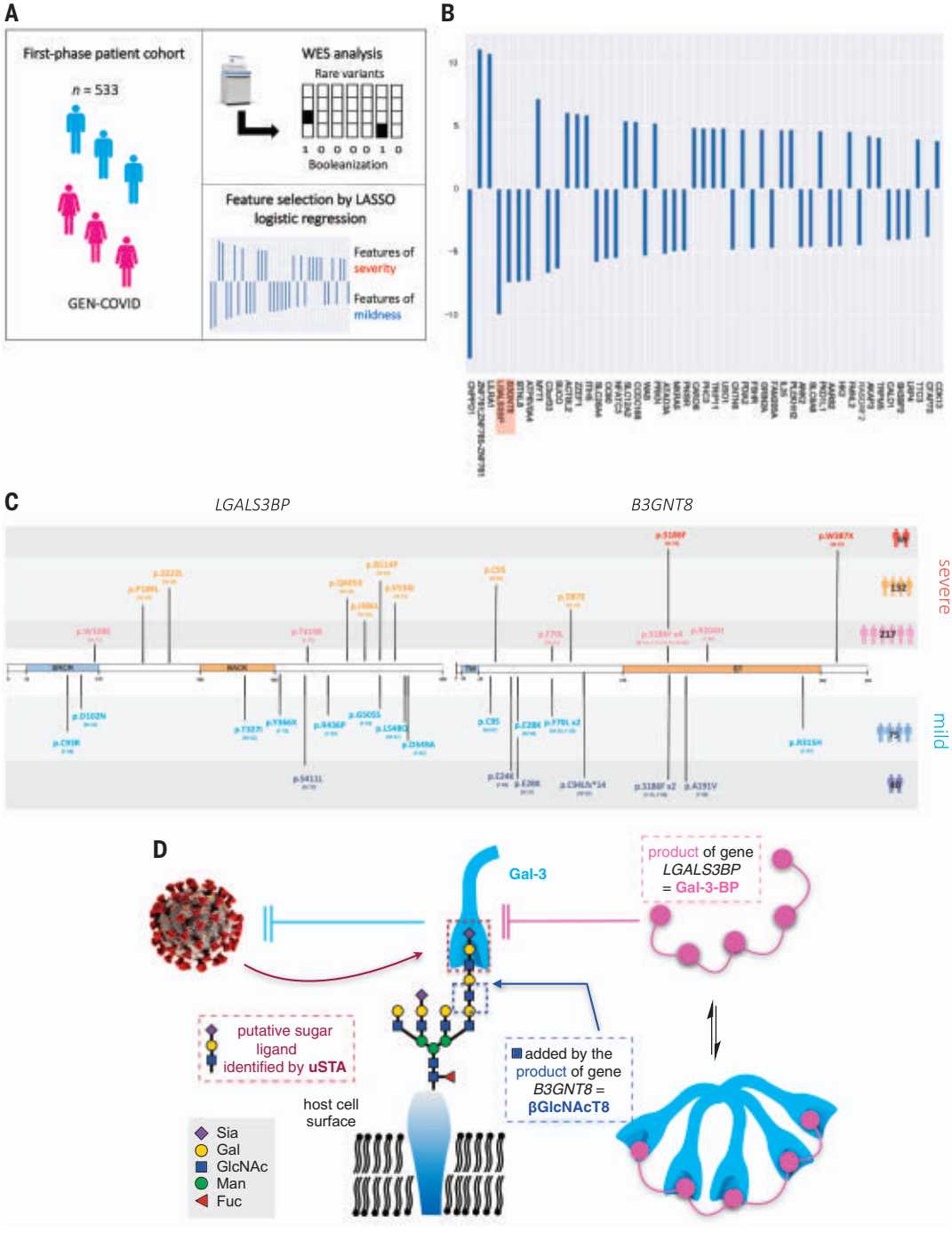


Fig. 7. Analyses of early 2020 first-phase SARS-CoV-2 PCR-positive patients reveals glycan-associated genes suggesting a model of glycan interaction consistent with uSTA observations of sialoside binding in B-origin-lineage SARS-CoV-2. (**A**) GEN-COVID workflow. Left: The GEN-COVID Multicenter Study cohort, of 533 SARS-CoV-2 PCR-positive subjects of different severity from phase one of the pandemic, was used for rare variant identification. Upper right: Whole-exome sequencing (WES) data were analyzed and binarized into 0 or 1 depending on the presence (1) or the absence (0) of variants in each gene. Lower right: LASSO logistic regression feature selection using a Boolean representation of WES data leads to the identification of final sets of features divided according to severity or mildness of disease, contributing to COVID-19 variability. See also (*79*, *80*) for further details of background methodology. (**B**) Histogram of the LASSO-based logistic regression weightings after recursive feature elimination analysis of 533 SARS-CoV-2–positive patients. Positive weights score susceptible response of gene variance to COVID-19 disease, whereas negative weights confer protective action through variance. Variation in glycan-associated genes *B3GNT8* and *LGALS3BP* score second and third out of all (>18,000) genes as the most protective, respectively (highlighted red). (**C**) Distribution of rare variants in *B3GNT8* and *LGALS3BP*. Left: Rare beneficial mutations distributed along the Gal-3-BP protein product of *LGALS3BP*, divided into the SRCR (scavenger receptor cysteine-rich) domain (light blue) and the BACK domain (light orange). Right: Rare beneficial mutations distributed along the βGlcNAcT8 protein product of *B3GNT8* divided into the predicted transmembrane (TM) domain (light blue) and glycosyltransferase catalytic (GT) domain (light orange), which catalyzes the transfer of polyLacNAc-initiating GlcNAc onto

tetraantennary *N*-linked glycoproteins [see also (D)]. The different colors of the mutation bands (top to bottom) refer to the severity grading of the PCR-positive patients who carried that specific mutation (red, Hospitalized intubated; orange, Hospitalized CPAP/BiPAP; pink, Hospitalized Oxygen Support; light blue, Hospitalized w/o Oxygen Support; blue, Not hospitalized a/pauci-symptomatic). (**D**) A proposed coherent model consistent with observation of implicated *B3GNT8* and *LGALS3BP* genes and the identification of sialosides as ligands for spike by uSTA and cryo-EM. Strikingly, although independently identified, *B3GNT8* and *LGALS3BP* produce gene products βGlcNAcT8 and Gal-3-BP, respectively, that manipulate and/or engage with processes associated with a common polyLacNAc-extended chain motif found on tetraantennary *N*-linked glycoproteins. A model emerges in which any associated loss of function from variance leads either to loss of polyLacNAc-extended chain (due to loss of initiation by βGlcNAcT8) or enhanced sequestration of by Gal-3 polyLacNAc-extended chain (which is antagonized by Gal-3-BP). Both would potentially lead to reduced access of virus spike to uSTA-identified motifs.

binding to certain sialosides was ablated in later phases of the pandemic (after September 2020) in variants of concern further highlights the dynamic role that sugar binding may play in virus evolution and may be linked, as has been previously suggested for H5N1 influenza A virus, to the "switching" of sugar-binding preferences by pathogens during or after zoonotic transitions (*1*). The focused "end-on" binding of the N-acetyl group in the N-acetyl-sialosides, which are found as the biosynthetically exclusive form of sialosides in humans (*60*), might have been a contributing factor in driving zoonosis.

Finally, our data also raise the question of why binding might be ablated in later variants of SARS-CoV-2. Again by comparison with influenza, which uses neuraminidases for the purpose of "release" when budding from a host cell (*61*), we speculate that in the absence of its own encoded neuraminidase, SARS-CoV-2 must walk a tight balance between the ability to bind human host glycans (potentially useful in a zoonotic leap) and cell-to-cell transmission (where release could become rate-limiting). One answer to this problem would be to ablate N-glycan binding via the sialoside motif subsequent to a successful zoonotic leap. This solution also has the advantage of removing a potential site for antibody neutralization for an interaction that might prove pivotal or critical in the context of zoonosis as a potentially global driver of virus fitness. Our combined data and models may therefore support decades-old hypotheses (*20*) proposing the benefit of cryptic sugar binding by pathogens that may be "switched on and off" to drive fitness in a different manner (e.g., in virulence or zoonosis) as needed.

Methods

Protein expression and purification: SARS-CoV-2 spike

The templates for wild type, alpha, and beta spike were kindly provided by P. Supasa and G. Screaton (University of Oxford). The gene encoding amino acids 1 to 1208 of the wild type, alpha, and beta SARS-CoV-2 spike glycoprotein ectodomain [with mutations of RRAR → GSAS at residues 682–685 (the furin cleavage site) and KV → PP at residues 986–987, as well as inclusion of a T4 fibritin trimerization domain] was cloned into the pOPINTTGneo-BAP vector using the forward primer (5′-GTCCAAGTTTATACTGAATTCCTCAAGCAGGCCACCATGTTCGTGTTCCTGGTGCTG-3′) and the reverse primer (5′-GTCATTCAGCAAGCTTAAAAAGGTAGAAAGTAATAC-3′), resulting in an aviTag/Bap sequence plus 6His in the 3′ terminus of the construct. The template for B-lineage-origin (wild type) spike is previously described (*62*). The templates for the alpha and beta spike are in the supplementary materials.

For B-lineage-origin, alpha, and beta spike, Expi293 cells (Thermofisher Scientific) were used to express the Spike-Bap protein. The cells were cultured in Expi293 expression media (Thermofisher Scientific) and were transfected using PEI MAX 40kDa (Polysciences) if cells were >95% viable and had reached a density of 1.5×10^6 to 2×10^6 cells per ml. Following transfection, cells were cultured at 37°C and 5% CO_2 at 120 rpm for 17 hours. Enhancers (6 mM valproic acid, 6.5 mM sodium propionate, 50 mM glucose, all from Sigma) were then added and protein was expressed at 30°C for 5 days before purification.

For delta and omicron spike, cDNA was synthesized (IDT) as gBlock, flanked by KpnI and XhoI restriction sites based on the HexaPro spike sequence (*63*). HexaPro delta spike was made based on the B-lineage-origin HexaPro spike with these additional mutations: T19R, G142D, E156G, del157/158, L452R, T478K, D614G, P681R, D950N. HexaPro Omicron BA.1 spike is made based on the original Wuhan HexaPro spike with these additional mutations: A67V, HV69-70 deletion, T95I, G142D, VYY143-145 deletion, N211 deletion, L212I, ins214EPE, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, F817P, N856K, A892P, A899P, A942P, Q954H, N969K, L981F, K986P, V987P. Mutations for delta and omicron spike were guided by the following databases: https://covariants.org/variants/21K.Omicron and https://viralzone.expasy.org/9556. The cDNA fragment was digested, cleaned up, and ligated to the pαH vector backbone using T4 Ligase (the same backbone as the original HexaPro Spike (B-lineage-origin) Addgene plasmid # 154754). Ligated plasmids were transformed in NEB DH5-alpha cells and plated on agar with ampicillin. Colonies were picked, cultured, and purified using the Qiagen HiSpeed MaxiPrep Kit and sent for Sanger sequencing to confirm the identity. Expi293F cells were transfected using ExpiFectamine293 transfection reagent (ThermoFisher) according to the manufacturer's instructions. The cells were cultured at 37°C, 8% CO_2, 125 rpm (25 mm throw) for 5 to 6 days before purification.

For the purification of wild type, alpha, beta, delta, and omicron spike, the medium in which the spike protein was secreted was supplemented with 1× PBS buffer at pH 7.4 (1:1 v/v) and 5 mM $NiSO_4$. The pH was adjusted with NaOH to pH 7.4 and filtered using a 0.8-μm filter. The mixture was stirred at 150 rpm for 2 hours at room temperature. The spike protein was purified on an Akta Express system (GE Healthcare) using a 5-ml His trap FF GE Healthcare column in PBS, 40 mM imidazole, pH 7.4, and eluted in PBS, 300 mM imidazole, pH 7.4. The protein was then injected onto either a Superdex 200 16/600 or 10/300 gel filtration column (GE Healthcare) in deuterated PBS buffer, pH 7.4. The eluted protein was concentrated using an Amicon Ultra-4 100kDa concentrator at 2000 rpm, 16°C (prewashed multiple times with deuterated PBS) to a concentration of roughly 1 mg/ml.

Protein expression and purification: Influenza HA

Freestyle 293-F cells were cultured in Freestyle expression media (Life Technologies) (37°C, 8% CO_2, 115 rpm orbital shaking). Cells were transfected at a density of 10^9 cells/liter with pre-incubated expression vector (300 μg/liter) and polyethyleneimine (PEI) MAX (Polysciences) (900 μg/liter). Expression vectors encoded terminally His-tagged wild-type influenza A virus (IAV) NC99 (H1N1) HA or a ΔRBS mutant previously described (*55*). After 5 days, supernatant was harvested and protein was purified via immobilized metal chromatography.

Protein expression and purification: C5 anti-spike nanobody

C5-Nanobody was purified as described (*46*). Purified C5 nanobody was then dialyzed into deuterated PBS buffer using 500 μl of Slide-A-Lyzer cassette (3.5 kDa cutoff).

Errors

The errors in the transfer efficiencies were estimated using a bootstrapping procedure. Specifically sample STD spectra were assembled through taking random combinations with replacement of mixing times, and the analysis to

obtain the transfer efficiency was performed on each. This process was repeated 100 times to enable evaluation of the mean and standard deviation transfer efficiency for each residue. Mean values correspond well with the value from the original analysis, and so we take the standard deviation as our estimate in uncertainty, which further is in accord with values obtained from independent repeated measurements.

Reagent sources

6′-Sialyllactose sodium salt and 3′-sialyllactose sodium salt were purchased from Carbosynth and used directly (6′-sialyllactose sodium salt, CAS-157574-76-0, 35890-39-2; 3′-sialyllactose sodium salt, CAS-128596-80-5, 35890-38-1). BSA and L-tryptophan were purchased from Sigma Aldrich. Heparin sodium salt, from porcine intestinal mucosa, IU ≥ 100/mg was purchased from Alfa Aesar. All other chemicals were purchased from commercial suppliers (Alfa Aesar, Acros, Sigma Aldrich, Merck, Carbosynth, Fisher, Fluorochem, VWR) and used as supplied, unless otherwise stated. See supplementary materials for syntheses of key compounds.

Protein NMR experiments

All NMR experiments in table S8 were conducted at 15°C on a Bruker AVANCE NEO 600 MHz spectrometer with CPRHe-QR-1H/19F/13C/15N-5mm-Z helium-cooled cryoprobe. Samples were stored in a Bruker SampleJet sample loader while not in magnet, at 4°C.

1D 1H NMR spectra with w5 water suppression were acquired using the Bruker pulse sequence zggpw5, using the smooth square Bruker shape SMSQ.10.100 for the pulsed-field gradients. The spectrum was centered on the water peak, and the receiver gain was adjusted. Typical acquisition parameters were sweep width of 9615.39 Hz, 16 scans per transient (NS), with four dummy scans, 32,768 complex points (TD), and a recycle delay (d1) of 1 s for a total acquisition time of 54 s. Reference 1D spectra of protein-only samples were acquired similarly with 16,384 scans per transient with a total acquisition time of 12.5 hours.

An STD experiment with excitation sculpted water suppression was developed from the Bruker pulse sequence stddiffesgp.2. The saturation was achieved using a concatenated series of 50-ms Gaussian-shaped pulses to achieve the desired total saturation time (d20). The shape of the pulses was specified by the Bruker shape file Gaus.1.1000, where the pulse is divided into 1000 steps and the standard deviation for the Gaussian shape is 165 steps. The field of the pulse was set to 200 Hz, which was calculated internally through scaling the power of the high-power 90° pulse.

The total relaxation delay was set to 5 s, during which the saturation pulse was applied. The data were acquired in an interleaved fashion, with each individual excitation frequency being repeated eight times (L4) until the total desired number of scans was achieved. Again, the spectrum was centered on the water peak, and the receiver gain was optimized. After recording of the free induction decay (FID), and prior to the recycle delay, a pair of water-selective pulses wee applied to destroy any unwanted magnetization. For all gradients (excitation sculpting and spoil), the duration was 3 ms using the smooth-square shape SMSQ10.100.

In a typical experiment, two excitation frequencies were required, one exciting protein, and one exciting far from the protein (+20,000 Hz, +33 ppm from the carrier). A range of mixing times were acquired to allow us to carefully quantify the buildup curve to obtain K_D values. A typical set of values used was 0.1 s, 0.3 s, 0.5 s, 0.7 s, 0.9 s, 1.1 s, 1.3 s, 1.5 s, 1.7 s, 1.9 s, 2.0 s, 2.5 s, 3.0 s, 3.5 s, 4.0 s, and 5.0 s.

Off- and on-resonance spectra were acquired for 16 saturation times, giving a total acquisition time of 8.7 hours.

The experiment was acquired as a pseudo-3D experiment, with each spectrum being acquired at a chosen set of excitation frequencies and mixing times. Recycle delays were set to 10 s for BSA + tryptophan STDs, and were 5 s otherwise.

For STD 10- to 50-ms Gaussian experiments, the saturation times used were every other time from the default STD: 0.1 s, 0.5 s, 0.9 s, 1.3 s, 1.7 s, 2 s, 3 s, 4 s.

List1: For STD var freq 1, the on-resonance frequencies in Hz relative to an offset of 2820.61 Hz are: 337.89, 422.36, 524.93, 736.11, 914.10, 1276.12, 1336.46, 1380.21, 1458.64, 1556.69, 1693.96, 2494.93, 2597.50, 2790.58, 2930.86, 3362.27, 3663.95, 3986.75, 4099.88, 4326.15, 4484.53, 4703.25, 4896.33, 5824.01, 6006.53, and 6208.65. The saturation times used were 2 s, 3 s, 4 s, and 5 s.

List2: For STD var freq 2, the on-resonance frequencies in Hz relative to an offset of 2820.61 Hz are: −2399.99, −1979.99, −1530.00, −1050.01, −330.021, 338.096, 1679.95, 1829.94, 1979.94, 2129.94, 2279.94, and 2579.93. The saturation times used were 0.1 s, 0.5 s, 2 s, and 5 s.

List3: For STD var freq 3, the on-resonance frequencies in Hz relative to an offset of 2820.61 Hz are: −2579.98, −2459.99, −2339.99, −2039.99, −1488.00, −1120.03, −345.02, 311.97, 1079.96, 1379.95, 1679.95, 1979.94, 2279.94, and 2579.93. The saturation times used were 0.1 s, 0.3 s, 0.5 s, and 0.9 s.

List4: For STD var freq 4, the on-resonance frequencies in Hz relative to an offset of 2820.61 Hz are: −2461.55, −1973.77, −1518.69, −1270.72, −693.69, −274.80, 280.08, 808.02, 1047.05, 2055.57, 2630.61, and 2979.65. The saturation times used were 0.1 s, 0.5 s, 0.9 s, 2 s, 3 s, and 4 s.

Spectra were also acquired on a 600-MHz spectrometer with Bruker Avance III HD console and 5-mm TCI CryoProbe, running TopSpin 3.2.6, recorded in table S9, and a 950-MHz spectrometer with Bruker Avance III HD console and 5-mm TCI CryoProbe, running TopSpin 3.6.1, recorded in table S11. The 950-MHz spectrometer used a SampleJet sample changer. Samples were stored at 15°C. The parameters used for the STD experiments were the same as above, with the following varying by instrument:

On the 600-MHz spectrometer, typical acquisition parameters were sweep width of 9615.39 Hz with typically 128 scans per transient (NS = 16 * L4 = 8), 32,768 complex points in the direct dimension and two dummy scans, executed prior to data acquisition.

On the 950-MHz spectrometer, typical acquisition parameters were sweep width of 15,243.90 Hz with typically 128 scans per transient (NS = 16 * L4 = 8), 32,768 complex points in the direct dimension and 2 dummy scans, executed prior to data acquisition.

uSTA data analysis

NMR spectra with a range of excitation frequencies and mixing times were acquired on ligand-only, protein-only, and mixed protein/ligand samples (fig. S6).

To analyze an STD dataset, two projections were created by summing over all 1D spectra and summing over all corresponding STD spectra. These two projections provide exceptionally high signal-to-noise, suitable for detailed analysis and reliable peak detection. The UnidecNMR algorithm was first executed on the 1D "pulse off" spectra to identify peak positions and intensities. Having identified possible peak positions, the algorithm then analyzes the STD spectra but only allowing resonances in places already identified in the 1D spectrum. Both analyses are conducted using the protein-only baselines for accurate effective subtraction of the protein baseline without the need to use relaxation filters (fig. S8).

The ligand-only spectra were analyzed similarly and in each case, excellent agreement with the known assignments was obtained, providing us with confidence in the algorithm. The mixed protein/ligand spectrum was then analyzed, which returned results very similar to the ligand-only case. Contributions from the protein, although small, were typically evident in the spectra, justifying the explicit inclusion of the protein-only baseline during the analysis. When analyzing the mixture, we included the protein-only background as a peak shape whose contribution to the spectrum could be freely adjusted. In this way, the spectra of protein/ligand mixtures could be accurately and quickly deconvolved, with the identified

ligand resonances occurring in precisely the positions expected from the ligand-only spectra. The results from the previous steps were then used to analyze the STD spectra. As these have much lower signal-to-noise, we fixed the ligand peak positions to be only those previously identified. Otherwise the protocol performed as described previously, where we used protein-only STD data to provide a baseline.

These analyses allow us to define a "transfer efficiency," which is simply the ratio of the signal from a given multiplet in the STD spectrum to the total expected in the raw 1D experiment. To obtain "per atom" transfer efficiencies, signals from the various pre-assumed components on the multiplets from each resonance were first summed before calculating the ratio. In the software, this is achieved by manually annotating the initial peak list using information obtained from independent-assignment experiments (see figs. S21 and S22).

Over the course of the project, it became clear that subtracting the transfer efficiencies obtained from a ligand-only sample was an essential part of the method (figs. S9 and S10). Depending on the precise relationship among the chemical shift of excitation, the location of the ligand peaks, and the excitation profile of the Gaussian train, we observed small apparent STD transfer in the ligand-only sample that cannot be attributed to ligand binding, arising from a small residual excitation of ligand protons, followed by internal cross-relaxation. It is likely that this excitation occurs at least in part via resonances of the ligand that are exchange-broadened, such as OH protons, which are not directly observed in the spectrum. When exciting far from the protein, zero ligand excitation is observed, as we would expect, but when exciting close to the methyls, or in the aromatic region, residual ligand excitation could be detected in ligand-only samples (figs. S9 and S10). Without the ligand-only correction, the uSTA surface may appear to be highly dependent on choice of excitation frequency. However, with the ligand correction, the relative uSTA profiles become invariant with excitation frequency. In general, therefore, we advise acquiring these routinely, and so the uSTA analysis assumes the presence of these data (figs. S6 and S8). The invariance of relative transfer efficiency with excitation frequency suggests that the internal evolution of magnetization within the protein during saturation (likely on the micro-/millisecond time scales) is much faster than the effective cross-relaxation rate between protein and ligand (occurring on the seconds time scale).

Having identified the relevant resonances of interest and performed both a protein and residual ligand subtraction, we reanalyzed the spectra without first summing over the different mixing times, in order to develop the quantitative atom-specific build-up curves. These were quantitatively analyzed as described below to obtain K_D and k_{off} rates. The values we obtain performing this analysis on BSA/Trp closely match those measured by ITC, and the values we measure for ligand **2** and Spike are in good agreement with those measured by SPR as described in the text.

The coverage of protons over the ligands studied here was variable; for example, there are no protons on carboxyl groups. To enable a complete surface to be rendered, the transfer efficiencies for each proton were calculated as described above, and the value is then transferred to the adjacent heteroatom. For heteroatoms not connected to an observed proton, a $1/r^6$ weighted average score was calculated. This approach allows us to define a unique surface. Caution should be exercised when quantitatively interpreting such surfaces where there are no direct measurements of the heteroatom.

In practice, raw unformatted FIDs are submitted to the uSTA pipeline, and the various steps described are performed largely automatically, where a user needs to manually adjust processing settings such as phasing and choosing which regions to focus on, iteratively adjust the peak shape to get a good match between the final reconvolved spectrum and the raw data, and input manual atomic assignments for each observed multiplet. The uSTA pipeline then provides a user with a report that shows the results of the various stages of analysis, and uses pymol to render the surfaces. The final transfer efficiencies delivered by the program can be combined with a folder containing a series of HADDOCK models to provide final structural models (Fig. 5).

Quantitative analysis via uSTA

In principle, a complete description of the saturation transfer experiment can be achieved via the Bloch-McConnell equations. If we can set up a density matrix describing all the spins in the system, their interactions, and their rates of chemical change in an evolution matrix **R**, then we can follow the system with time according to:

$$\rho(t) = \rho(0)\exp(-Rt)$$

The challenge comes from the number of spins that must be included and the need to accurately describe all the interactions between them, which will need to also include how these are modulated by molecular motions in order to get an accurate description of the relaxation processes. This is illustrated by the CORCEMA method (*64*) that takes a static structure of a protein/ligand complex and estimates STD transfers. The CORCEMA calculations performed to arrive at cross-relaxation rates assume the complex is rigid, which is often a poor approximation for a protein, and because of the large number of spins involved, the calculation is sufficiently intensive such that this calculation cannot be routinely used to fit to experimental data.

It would be very desirable to extract quantitative structural parameters, as well as chemical properties such as interaction strengths and association/dissociation rates, directly from STD data. In what follows, we develop a simple quantitative model for the STD experiment to achieve this goal. We will treat the system as comprising just two spins, one to represent the ligand and one to represent the protein, and we allow the two spins to exist either in isolation or in a bound state. We can safely neglect scalar coupling and so we only need to allow the x, y, and z basis operators for each spin, together with an identity operator to ensure that the system returns to thermal equilibrium at long times. As such, our evolution matrix **R** will be a square matrix with 13 × 13 elements.

For the spin part, our model requires us to consider the chemical shift of the ligand in the free and bound states, and the chemical shifts of the protein in the free and bound states. In practice however, it is sufficient to set the free protein state on resonance with the pulse, and the free ligand chemical shift is set to a value that matches experiment.

The longitudinal and transverse relaxation rates are calculated for the free and bound states using a simple model assuming in each state there are two dipole-coupled spins separated by a distance R with similar Larmor frequency. In addition, cross-relaxation between ligand and protein is allowed only when the two are bound. The relaxation rates are characterized by an effective distance and an effective correlation time,

$$R_1 = \frac{1}{4}K[J(0) + 3J(\omega) + 6J(2\omega)]$$

$$R_2 = \frac{1}{4}K\left[\frac{5}{2}J(0) + \frac{9}{2}J(\omega) + 3J(2\omega)\right]$$

$$\sigma = \frac{1}{4}K[6J(2\omega) - J(0)]$$

which are each parameterized in terms of an interaction constant (depending on effective distance)

$$K = \left(\frac{\mu_0 \hbar \gamma_H^2}{4\pi r^3}\right)^2$$

and a spectral density function (depending on effective distance)

$$J(\omega) = \frac{2}{5}\frac{\tau}{1 + \omega^2\tau^2}$$

The longitudinal and transverse relaxation rates R_1 and R_2 describe auto-relaxation of diagonal z elements and xy elements, respectively. The cross-relaxation rates σ describe cross-relaxation and couple z elements between the ligand and protein in the bound state. We ensure that the system returns to equilibrium at long times by

adding elements of the form R_1M_0 or σM_0 linking the identify element and the z matrix elements. Overall, the relaxation part of the model is parameterized by two correlation times, one for the ligand and one for the protein/complex, and three distances, one for the ligand auto-relaxation rates, one for protein auto-relaxation rates, and one for the protein/ligand separation.

Finally, the chemical kinetics govern the rates at which the spins can interconvert. We will take a simple model where $PL \leftrightarrows P + L$, whose dissociation constant is given by

$$K_D = \frac{k_{off}}{k_{on}} = \frac{[P][L]}{[PL]}$$

The free protein concentration can be determined from knowledge of the K_D and the total ligand and protein concentrations:

$$[P] = \frac{1}{2}\left[P_{Tot} - L_{Tot} - K_D + \sqrt{(L_{Tot} + K_D - P_{Tot})^2 + 4P_{Tot}K_D}\right]$$

from which the bound protein concentration and the free and bound ligand concentrations can be easily calculated.

The density matrix is initialized with the free and bound protein/ligand concentrations assigned to the relevant z operators. It was found to be important to additionally include a factor that accounts for the increased proton density within the protein. The saturation pulse is then applied either as a concatenated series of Gaussian pulses whose duration and peak power in Hz needs to be specified, exactly matching the pulse shapes and durations used in the experiment (see NMR methods above).

Build-up curves and transfer efficiencies can be easily simulated using this model and compared to data, and the various parameters can be optimized to fit to the data. In total, the model is characterized by eight parameters: K_D, k_{off}, the correlation times of the ligand and the protein, the three distances described above, and the proton density within the protein. There is substantial correlation between the effects of the various parameters, and care is needed using optimization to avoid local minima. By obtaining data at various protein and ligand concentrations, however, it is possible to break this degeneracy and obtain well-described values as in the text.

In practical terms, the initial rate of the build-up curve is predominantly affected by the cross-relaxation rate and the off rate, and the final height of the build-up curve is mostly influenced by the proton density in the protein and K_D. Software to perform this analysis has been directly incorporated into the uSTA software.

Parameters fitted by the model

Overall the model is parameterized by a set of values that characterize the intrinsic and cross relaxation. From τ_G and r_{IS}(ligand) we estimate R_1 and R_2 of the ligand; from τ_E and r_{IS}(protein) we obtain R_1 and R_2 of the protein; and from τ_E and r_{IS}(complex) we calculate the cross-relaxation rate. These values are combined with a factor that accounts for the larger number of spins present in the protein, "fac," and the on and off rates, to complete a set of eight parameters that specify our model. The distances should be considered "effective" values that parameterize the relaxation rates, although in principle it should be possible to obtain physical insights from their interpretation. The concentration-independent relaxation rates can be separated from the exchange rates by comparing the curves as a function of ligand and protein concentration. By treating the system as comprising two spins, we are effectively assuming that the cross-relaxation within the protein is very efficient. In the STD experiment, saturation pulses are applied for several seconds, which is sufficient for near-saturating spin diffusion within a protein. Because of the complexity of the model, optimization of the parameters via a gradient descent method can get stuck in local minima. In practical applications, it is advisable to start the optimization over a range of initial conditions, particularly in the rates, to ensure that the lowest possible χ^2 is achieved.

Thermostability assays

Thermal stability assays were performed using a NanoTemper Prometheus NT.48 (Membrane Protein Laboratory, Diamond Light Source). To 11 μl of 2 μM spike (deuterated PBS), 2 μl of trisaccharide 2 (deuterated PBS) was titrated to give final concentrations of 0.1, 0.2, 0.4, 0.8, 1.6, and 2.0 mM. Samples were then loaded into capillaries and heated from 15° to 95°C. Analysis was performed using PR.ThermControl v2.3.1 software.

SPR binding measurement assays

All experiments were performed on a Biacore T200 instrument. For the immobilization of SiaLac onto the sensor chip, a flow rate of 10 μl/min was used in a buffer solution of HBP-EP (0.01 M HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005% v/v surfactant P20). A CM5 sensor chip (carboxymethylated dextran) was equilibrated with HBS-EP buffer at 20°C. The chip was activated by injecting a mixture of *N*-hydroxysuccinimide (50 mM) and EDC-HCl (200 mM) for 10 min followed by a 2-min wash step with buffer. Ethylenediamine (1 M in PBS) was then injected for 7 min followed by a 2-min wash step followed by ethanolamine-HCl (1 M, pH 8.5) for 10 min and then a further 1-min wash step. Finally, SiaLac-IME (5.6 mM in PBS) reagent **10** was injected over 10 min and a final 2-min wash step was performed (see fig. S13 and supplementary materials for further details).

For analysis of spike binding, a flow rate of 10 μl/min was used at 16°C. Serial dilutions of spike (0.19, 0.50, 1.36, and 3.68 μM) were injected for 30 s association and 150 s dissociation starting with the lowest concentration. Buffer-only runs were carried out before injection of spike and after the first two dilutions. BSA (3.03 μM in PBS) was used as a negative control, and a mouse serum in a 100-fold dilution was used as a positive control.

Analysis of SPR data

To analyze the SPR data, we assume an equilibrium of the form $PL \leftrightarrows P + L$ characterized by a dissociation constant

$$K_D = \frac{k_{off}}{k_{on}} = \frac{[P][L]}{[PL]}$$

To follow the kinetics of binding and dissociation, we assume that the SPR response Γ is proportional to the bound complex $\Gamma = \kappa[PL]$, which leads to the following kinetic equation:

$$\frac{d\Gamma}{dt} + \frac{k_{off}}{\kappa}\Gamma - \frac{k_{on}}{\kappa}[P][L] = 0$$

This can be solved when restrained by the total number of binding sites, $L_{tot} = [L] + [PL]$. Under conditions of constant flow, we assume that the free protein concentration is constant, which leads to the following:

$$\Gamma_{on} = \frac{\kappa k_{on} L_{tot}[P]}{k_{off} + k_{on}[P]}\{1 - \exp[-(k_{off} + k_{on}[P])t]\}$$

And similarly, for dissociation where we take the concentration of free protein to be zero:

$$\Gamma_{off} = \kappa \mathrm{Re}^{-k_{off}t}$$

The recovery of the chip was not complete after each protein concentration and wash step, as has been observed for shear-induced lectin-ligand binding with glycans immobilized onto a chip surface (*65*). Nonetheless, the data were well explained by a global analysis where the on and off rates were held to be identical for each replicate, but the value of k was allowed to vary slightly between runs, and an additional constant was introduced to Γ_{off} to account for incomplete recovery of the SPR signal following standard approaches. Concentration of spike was insufficient to get the plateau region of the binding, and so the specific time values taken for the on rate affect the fitted values.

Modeling of the N-terminal domain of SARS-CoV-2 with glycans

We modeled the structure of the NTD on Protein Data Bank (PDB) entry 7c2l (*19*) because it provided much better coverage of the area of interest when compared to the majority of the templates available at PDB as of 15 July 2020. The models were created with Modeller (*66*),

using the "automodel" protocol without refining the "loop." We generated 10 models and ranked them by their DOPE score (*67*), selecting the top five for ensemble docking.

Docking of 3′-sialyllactose to SARS-CoV-2 NTD

We docked 3′-sialyllactose to NTD with version 2.4 of the HADDOCK webserver (*42*, *43*). The binding site on NTD was defined by comparison with PDB entry 6q06 (*5*), a complex of MERS-CoV spike protein and 2,3-sialyl-*N*-acetyl-lactosamine. The binding site could not be directly mapped because of conformational differences between the NTDs of MERS-CoV and SARS-CoV-2, but by inspection a region with similar properties (aromatics, methyl groups, and positively charged residues) could be identified. We defined in HADDOCK the sialic acid as "active" and residues 18, 19, 20, 21, 22, 68, 76, 77, 78, 79, 244, 254, 255, 256, 258, and 259 of NTD as "passive," meaning the sialic acid needs to make contact with at least one of the NTD residues but there is no penalty if it doesn't contact all of them, thus allowing the compound to freely explore the binding pocket. Because only one restraint was used, we disabled the random removal of restraints. Following our small-molecule docking recommended settings (*68*), we skipped the "hot" parts of the semi-flexible simulated annealing protocol ("initiosteps" and "cool1_steps" set to 0) and also lowered the starting temperature of the last two substages to 500 and 300 K, respectively ("tadinit2_t" and "tadinit3_t" to 500 and 300, respectively). Clustering was performed based on "RMSD" with a distance cutoff of 2 Å, and the scoring function was modified to

$$\begin{aligned}\text{HADDOCKscore} = {} & 1.0{*}E_{\text{vdW}} + 0.1{*}E_{\text{elec}} \\ & + 1.0{*}E_{\text{desol}} + 0.1{*}E_{\text{AIR}}\end{aligned}$$

All other settings were kept to their default values. Finally, the atom-specific transfer efficiencies determined by uSTA were used to filter cluster candidates.

Cryo-EM analysis

SARS-CoV-2 spike protein, generated and purified as described (*48*), in 1.1 mg/ml was incubated with 10 mM ethyl(triiodobenzamide) siallyllactoside overnight at 4°C. A 3.5-μl sample was applied to glow-discharged Quantifoil gold R1.2/1.3 300-mesh grids and blotted for ~3 s at 100% humidity and 6°C before vitrification in liquid ethane using Vitrobot (FEI). Two datasets were collected on Titan Krios equipped with a K2 direct electron detector at the cryo-EM facility (OPIC) in the Division of Structural Biology, University of Oxford. Both datasets were collected by SerialEM at a magnification of 165,000× with a physical pixel size of 0.82 Å per pixel. Defocus range was −0.8 μm to −2.4 μm. Total doses for the two datasets were 60 e/Å^2 (5492 movies) and 61 e/Å^2 (8284 movies), respectively.

Motion correction was performed by MotionCor2 (*69*). The motion-corrected micrographs were imported into cryoSPARC (*70*) and contrast transfer function values were estimated using Gctf (*71*) in cryoSPARC. Templates were produced by 2D classification from 5492 micrographs with particles auto-picked by Laplacian-of-Gaussian (LoG)–based algorithm in RELION 3.0 (*72*, *73*). Particles were picked from all micrographs using Template picker in cryoSPARC. Multiple rounds of 2D classification were carried out and the selected 2D classes (372,157 particles) were subjected to 3D classification (Heterogeneous Refinement in cryoSPARC) using six classes. One class was predominant after 3D classification. Nonuniform refinement (*74*) was performed for this class (312,018 particles) with *C*1 and *C*3 symmetry, respectively, yielding a 2.27 Å map for *C*3 symmetry and a 2.44 Å map for *C*1 symmetry. See also fig. S23 and table S12 for cryo-EM data collection, refinement, and validation statistics.

Genetic analysis of clinical samples

Variant calling: Reads were mapped to the hg19 reference genome by the Burrow-Wheeler aligner BWA. Variants calling was performed according to the GATK4 best practice guidelines. Namely, duplicates were first removed by *MarkDuplicates*, and base qualities were recalibrated using *BaseRecalibration* and *ApplyBQSR*. *HaplotypeCaller* was used to calculate Genomic VCF files for each sample, which were then used for multi-sample calling by *GenomicDBImport* and *GenotypeGVCF*. In order to improve the specificity-sensitivity balance, variants' quality scores were calculated by *VariantRecalibrator* and *ApplyVQSR*, and only variants with estimated truth sensitivity above 99.9% were retained. Variants were annotated by ANNOVAR.

Rare variant selection: Missense, splicing, and loss-of-function variants with a frequency lower than 0.01 according to ExAC_NFE (Non Finnish European ExAC Database) were considered for further analyses. A score of 0 was assigned to each sample where the gene is not mutated, and a score of 1 was assigned when at least one variant is present on the gene.

The cohort was distributed as follows. Ethnicity: 504 white, 4 Black, 5 Asian, 16 Hispanic ethnicity, 4 patients for which this information was not available. Sex: 317 male, 216 female. Age: minimum age 19 years, maximum age 99 years, mean age 62.5 years.

Gene prioritization by logistic regression

Discriminating genes in COVID-19 disease were interpreted in a framework of feature selection analysis using a customized feature selection approach based on the recursive feature elimination algorithm applied to the LASSO logistic regression model. Specifically, for a set of n samples $\{x_i, y_i\}$ $(i = 1, \ldots, n)$, each of which consists of p input features $x_{i,k} \in \chi_i$, $k = 1, \ldots, p$ and one output variable $y_i \in Y$, these features assumed the meaning of genes, whereas the samples were the patients involved in the study. The space $\chi = \chi_1 \times \chi_2 \ldots \times \chi_p$ was denoted "input space," whereas the "hypothesis space" was the space of all the possible functions f: $\chi \rightarrow Y$ mapping the inputs to the output. Given that the number of features (p) is substantially higher than the number of samples (n), LASSO regularization (*49*) has the effect of shrinking the estimated coefficients to zero, providing a feature selection method for sparse solutions within the classification tasks. Feature selection methods based on such regularization structures (embedded methods) were most applicable to our scope because they were computationally tractable and strictly connected with the classification task of the ML algorithm.

As the baseline algorithm for the embedded method, we adopted the logistic regression (LR) model that is a state-of-the-art ML algorithm for binary classification tasks with probabilistic interpretation. It models the log-odds of the posterior success probability of a binary variable as the linear combination of the input:

$$\log\frac{\Pr(Y=1|X=\mathbf{x})}{1-\Pr(Y=1|X=\mathbf{x})} = \beta_0 + \sum_{k=1}^{p}\beta_k x_k$$

where $\mathbf{x}$ is the input vector, β_k are the coefficients of the regression, and X and Y are the random variables representing the input and the output, respectively. The loss function to be minimized is given by the binary cross-entropy loss

$$-\sum_{i=1}^{n}[y_i \log \hat{y}_i (1-y_i)\log(1-\hat{y}_i)]$$

where $\hat{y} = \Pr(Y = 1|X = \mathbf{x})$ is the predicted target variable and y is the true label. As already introduced, in order to enforce both the sparsity and the interpretability of the results, the model is trained with the additional LASSO regularization term

$$\lambda\sum_{k=1}^{p}|\beta_k|$$

In this way, the absolute value of the surviving weights of the LR algorithm was interpreted as the feature importance of the subset of most relevant genes for the task. Because a feature-ranking criterion can become suboptimal when the subset of removed features is large (*75*), we applied recursive feature elimination (RFE) methodology. For each step of the procedure, we fitted the model and removed the features with smallest ranking criteria in a recursive manner until a certain number of features was reached.

The fundamental hyperparameter of LR is the strength of the LASSO term tuned with a grid search procedure on the accuracy of the 10-fold cross-validation. The k-fold cross-validation provided the partition of the dataset into k batches, then exploited k^{-1} batches for the training and the remaining test batch as a test, by repeating this procedure k times. In the grid search method, a cross-validation procedure was carried out for each value of the regularization hyperparameter in the range $[10^{-4}, ..., 10^{6}]$. Specifically, the optimal regularization parameter is chosen by selecting the most parsimonious parameter whose cross-validation average accuracy falls in the range of the best one along with its standard deviation. During the fitting procedure, the class unbalancing was tackled by penalizing the misclassification of minority class with a multiplicative factor inversely proportional to the class frequencies. For the RFE, the number of included features at each step of the algorithm as well as the final number of features was fixed at 100. All data preprocessing and the RFE procedure were coded in Python; the LR model was used, as included, in the scikit-learn module with the liblinear coordinate descent optimization algorithm.

REFERENCES AND NOTES

1. S. Yamada *et al*., Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. *Nature* **444**, 378–382 (2006). doi: 10.1038/nature05264; pmid: 17108965
2. W. Li *et al*., Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E8508–E8517 (2017). doi: 10.1073/pnas.1712592114; pmid: 28923942
3. C. Schwegmann-Wessels, G. Herrler, Sialic acids as receptor determinants for coronaviruses. *Glycoconj. J.* **23**, 51–58 (2006). doi: 10.1007/s10719-006-5437-9; pmid: 16575522
4. R. J. G. Hulswit *et al*., Human coronaviruses OC43 and HKU1 bind to 9-*O*-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2681–2690 (2019). doi: 10.1073/pnas.1809667116; pmid: 30679277
5. Y.-J. Park *et al*., Structures of MERS-CoV spike glycoprotein in complex with sialoside attachment receptors. *Nat. Struct. Mol. Biol.* **26**, 1151–1157 (2019). doi: 10.1038/s41594-019-0334-7; pmid: 31792450
6. E. Qing, M. Hantak, S. Perlman, T. Gallagher, Distinct Roles for Sialoside and Protein Receptors in Coronavirus Infection. *mBio* **11**, e02764-19 (2020). doi: 10.1128/mBio.02764-19; pmid: 32047128
7. W. Li *et al*., Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454 (2003). doi: 10.1038/nature02145; pmid: 14647384
8. K. Kuba *et al*., A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat. Med.* **11**, 875–879 (2005). doi: 10.1038/nm1267; pmid: 16007097
9. F. Li, W. Li, M. Farzan, S. C. Harrison, Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864–1868 (2005). doi: 10.1126/science.1116480; pmid: 16166518
10. D. Wrapp *et al*., Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020). doi: 10.1126/science.abb2507; pmid: 32075877
11. R. Yan *et al*., Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444–1448 (2020). doi: 10.1126/science.abb2762; pmid: 32132184
12. D. Morniroli, M. L. Giannì, A. Consales, C. Pietrasanta, F. Mosca, Human Sialome and Coronavirus Disease-2019 (COVID-19) Pandemic: An Understated Correlation? *Front. Immunol.* **11**, 1480 (2020). doi: 10.3389/fimmu.2020.01480; pmid: 32655580
13. A. N. Baker *et al*., The SARS-COV-2 Spike Protein Binds Sialic Acids and Enables Rapid Detection in a Lateral Flow Point of Care Diagnostic Device. *ACS Cent. Sci.* **6**, 2046–2052 (2020). doi: 10.1021/acscentsci.0c00855; pmid: 33269329
14. W. Hao *et al*., Binding of the SARS-CoV-2 spike protein to glycans. *Sci. Bull.* **66**, 1205–1214 (2021). doi: 10.1016/j.scib.2021.01.010; pmid: 33495714
15. L. Nguyen *et al*., Sialic acid-containing glycolipids mediate binding and viral entry of SARS-CoV-2. *Nat. Chem. Biol.* **18**, 81–90 (2022). doi: 10.1038/s41589-021-00924-1
16. W. Hao *et al*., Binding of the SARS-CoV-2 Spike Protein to Glycans. bioRxiv 2020.2005.2017.100537 (2020). doi: 10.1101/2020.05.17.100537
17. T. M. Clausen *et al*., SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. *Cell* **183**, 1043–1057.e15 (2020). doi: 10.1016/j.cell.2020.09.033; pmid: 32970989
18. N. Behloul, S. Baha, R. Shi, J. Meng, Role of the GTNGTKR motif in the N-terminal receptor-binding domain of the SARS-CoV-2 spike protein. *Virus Res.* **286**, 198058 (2020). doi: 10.1016/j.virusres.2020.198058; pmid: 32531235
19. X. Chi *et al*., A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* **369**, 650–655 (2020). doi: 10.1126/science.abc6952; pmid: 32571838
20. M. G. Rossmann, The canyon hypothesis. *J. Biol. Chem.* **264**, 14587–14590 (1989). doi: 10.1016/S0021-9258(18)63732-9; pmid: 2670920
21. Y. Watanabe *et al*., Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nat. Commun.* **11**, 2688 (2020). doi: 10.1038/s41467-020-16567-0; pmid: 32461612
22. M. Mayer, B. Meyer, Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chem. Int. Ed.* **38**, 1784–1788 (1999). doi: 10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q; pmid: 29711196
23. J. L. Wagstaff, S. L. Taylor, M. J. Howard, Recent developments and applications of saturation transfer difference nuclear magnetic resonance (STD NMR) spectroscopy. *Mol. Biosyst.* **9**, 571–577 (2013). doi: 10.1039/C2MB25395J; pmid: 23232937
24. J. Angulo, P. M. Enríquez-Navas, P. M. Nieto, Ligand-receptor binding affinities from saturation transfer difference (STD) NMR spectroscopy: The binding isotherm of STD initial growth rates. *Chem. Eur. J.* **16**, 7803–7812 (2010). doi: 10.1002/chem.200903528; pmid: 20496354
25. R. H. McMenamy, J. L. Oncley, The specific binding of L-tryptophan to serum albumin. *J. Biol. Chem.* **233**, 1436–1447 (1958). doi: 10.1016/S0021-9258(18)49353-2; pmid: 13610854
26. P. Vallurupalli, G. Bouvignies, L. E. Kay, Studying "invisible" excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.* **134**, 8148–8161 (2012). doi: 10.1021/ja3001419; pmid: 22554188
27. T. Xie, T. Saleh, P. Rossi, C. G. Kalodimos, Conformational states dynamically populated by a kinase determine its function. *Science* **370**, eabc2754 (2020). doi: 10.1126/science.abc2754; pmid: 33004676
28. N. L. Fawzi, J. Ying, D. A. Torchia, G. M. Clore, Probing exchange kinetics and atomic resolution dynamics in high-molecular-weight complexes using dark-state exchange saturation transfer NMR spectroscopy. *Nat. Protoc.* **7**, 1523–1533 (2012). doi: 10.1038/nprot.2012.077; pmid: 22814391
29. G. Bouvignies *et al*., Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **477**, 111–114 (2011). doi: 10.1126/science.abc2754; pmid: 33004676
30. M. Mayer, B. Meyer, Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.* **123**, 6108–6117 (2001). doi: 10.1021/ja0100120; pmid: 11414845
31. M. T. Marty *et al*., Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370–4376 (2015). doi: 10.1021/acs.analchem.5b00140; pmid: 25799115
32. L. Fielding, S. Rutherford, D. Fletcher, Determination of protein-ligand binding affinity by NMR: Observations from serum albumin model systems. *Magn. Reson. Chem.* **43**, 463–470 (2005). doi: 10.1002/mrc.1574; pmid: 15816062
33. A. Bujacz, K. Zielinski, B. Sekula, Structural studies of bovine, equine, and leporine serum albumin complexes with naproxen. *Proteins* **82**, 2199–2208 (2014). doi: 10.1002/prot.24583; pmid: 24753230
34. I. Pérez-Victoria *et al*., Saturation transfer difference NMR reveals functionally essential kinetic differences for a sugar-binding repressor protein. *Chem. Commun.* **2009**, 5862–5864 (2009). doi: 10.1039/b913489a; pmid: 19787122
35. U. Hars, R. Horlacher, W. Boos, W. Welte, K. Diederichs, Crystal structure of the effector-binding domain of the trehalose-repressor of Escherichia coli, a member of the LacI family, in its complexes with inducer trehalose-6-phosphate and noninducer trehalose. *Protein Sci.* **7**, 2511–2521 (1998). doi: 10.1002/pro.5560071204; pmid: 9865945
36. H. M. McConnell, Reaction Rates by Nuclear Magnetic Resonance. *J. Chem. Phys.* **28**, 430–431 (1958). doi: 10.1063/1.1744152
37. J. E. Stencel-Baerenwald, K. Reiss, D. M. Reiter, T. Stehle, T. S. Dermody, The sweet spot: Defining virus-sialic acid interactions. *Nat. Rev. Microbiol.* **12**, 739–749 (2014). doi: 10.1038/nrmicro3346; pmid: 25263223
38. D. Lingwood *et al*., Structural and genetic basis for development of broadly neutralizing influenza antibodies. *Nature* **489**, 566–570 (2012). doi: 10.1038/nature11371; pmid: 22932267
39. H.-Y. Liao *et al*., Differential receptor binding affinities of influenza hemagglutinins on glycan arrays. *J. Am. Chem. Soc.* **132**, 14849–14856 (2010). doi: 10.1021/ja104657b; pmid: 20882975
40. Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan, M. Crispin, Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **369**, 330–333 (2020). doi: 10.1126/science.abb9983; pmid: 32366695
41. M. P. Lenza *et al*., Structural Characterization of N-Linked Glycans in the Receptor Binding Domain of the SARS-CoV-2 Spike Protein and their Interactions with Human Lectins. *Angew. Chem. Int. Ed.* **59**, 23763–23771 (2020). doi: 10.1002/anie.202011015; pmid: 32915505
42. C. Dominguez, R. Boelens, A. M. J. J. Bonvin, HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737 (2003). doi: 10.1021/ja026939x; pmid: 12580598
43. G. C. P. van Zundert *et al*., The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **428**, 720–725 (2016). doi: 10.1016/j.jmb.2015.09.014; pmid: 26410586
44. A. Rambaut *et al*., A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020). doi: 10.1038/s41564-020-0770-5; pmid: 32669681
45. J. Huo *et al*., Neutralizing nanobodies bind SARS-CoV-2 spike RBD and block interaction with ACE2. *Nat. Struct. Mol. Biol.* **27**, 846–854 (2020). doi: 10.1038/s41594-020-0469-6; pmid: 32661423
46. J. Huo *et al*., A potent SARS-CoV-2 neutralising nanobody shows therapeutic efficacy in the Syrian golden hamster model of COVID-19. *Nat. Commun.* **12**, 5469 (2021). doi: 10.1038/s41467-021-25480-z; pmid: 34552091
47. M. McCallum, A. C. Walls, J. E. Bowen, D. Corti, D. Veesler, Structure-guided covalent stabilization of coronavirus spike glycoprotein trimers in the closed conformation. *Nat. Struct. Mol. Biol.* **27**, 942–949 (2020). doi: 10.1038/s41594-020-0483-8; pmid: 32753755
48. C. Toelzer *et al*., Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science* **370**, 725–730 (2020). doi: 10.1126/science.abd3255; pmid: 32958580
49. R. Tibshirani, Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. B* **58**, 267–288 (1996). doi: 10.1111/j.2517-6161.1996.tb02080.x
50. K. Koths, E. Taylor, R. Halenbeck, C. Casipit, A. Wang, Cloning and characterization of a human Mac-2-binding protein, a new member of the superfamily defined by the macrophage scavenger receptor cysteine-rich domain. *J. Biol. Chem.* **268**, 14245–14249 (1993). doi: 10.1016/S0021-9258(19)85233-X; pmid: 8390986
51. L. Johannes, R. Jacob, H. Leffler, Galectins at a glance. *J. Cell Sci.* **131**, jcs208884 (2018). doi: 10.1242/jcs.208884; pmid: 29717004
52. S. R. Stowell *et al*., Galectin-1, -2, and -3 exhibit differential recognition of sialylated glycans and blood group antigens. *J. Biol. Chem.* **283**, 10109–10123 (2008). doi: 10.1074/jbc.M709545200; pmid: 18216021
53. N. A. Kamili *et al*., Key regulators of galectin-glycan interactions. *Proteomics* **16**, 3111–3125 (2016). doi: 10.1002/pmic.201600116; pmid: 27582340
54. A. Togayachi *et al*., β3GnT2 (*B3GNT2*), a major polylactosamine synthase: Analysis of B3GNT2-deficient mice. *Methods Enzymol.* **479**, 185–204 (2010). doi: 0.1016/s0076-6879(10)79011-x
55. M. Kanekiyo *et al*., Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies.

Nature **499**, 102–106 (2013). doi: 10.1038/nature12202; pmid: 23698367
56. M. Mammen, S.-K. Choi, G. M. Whitesides, Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *Angew. Chem. Int. Ed.* **37**, 2754–2794 (1998). doi: 10.1002/(SICI)1521-3773(19981102)37:20<2754::AID-ANIE2754>3.0.CO;2-3; pmid: 29711117
57. R. T. Lee, Y. C. Lee, Affinity enhancement by multivalent lectin-carbohydrate interaction. *Glycoconj. J.* **17**, 543–551 (2000). doi: 10.1023/A:1011070425430; pmid: 11421347
58. N. Jia *et al*., The Human Lung Glycome Reveals Novel Glycan Ligands for Influenza A Virus. *Sci. Rep.* **10**, 5320 (2020). doi: 10.1038/s41598-020-62074-z; pmid: 32210305
59. M. A. Toscano *et al*., Differential glycosylation of TH1, TH2 and TH-17 effector cells selectively regulates susceptibility to cell death. *Nat. Immunol.* **8**, 825–834 (2007). doi: 10.1038/ni1482; pmid: 17589510
60. P. Tangvoranuntakul *et al*., Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12045–12050 (2003). doi: 10.1073/pnas.2131556100; pmid: 14523234
61. J. L. McAuley, B. P. Gilbertson, S. Trifkovic, L. E. Brown, J. L. McKimm-Breschkin, Influenza Virus Neuraminidase Structure and Functions. *Front. Microbiol.* **10**, 39 (2019). doi: 10.3389/fmicb.2019.00039; pmid: 30761095
62. J. Huo *et al*., Neutralization of SARS-CoV-2 by Destruction of the Prefusion Spike. *Cell Host Microbe* **28**, 445–454.e6 (2020). doi: 10.1016/j.chom.2020.06.010; pmid: 32585135
63. C.-L. Hsieh *et al*., Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* **369**, 1501–1505 (2020). doi: 10.1126/science.abd0826; pmid: 32703906
64. V. Jayalakshmi, N. R. Krishna, Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes. *J. Magn. Reson.* **155**, 106–118 (2002). doi: 10.1006/jmre.2001.2499; pmid: 11945039
65. K. Nakamura *et al*., Immobilized glycosylated Fmoc-amino acid for SPR: Comparative studies of lectin-binding to linear or biantennary diLacNAc structures. *Carbohydr. Res.* **382**, 77–85 (2013). doi: 10.1016/j.carres.2013.10.003; pmid: 24211369
66. A. Sali, Comparative protein modeling by satisfaction of spatial restraints. *Mol. Med. Today* **1**, 270–277 (1995). doi: 10.1016/S1357-4310(95)91170-7; pmid: 9415161
67. M. Y. Shen, A. Sali, Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507–2524 (2006). doi: 10.1110/ps.062416606; pmid: 17075131
68. P. I. Koukos, L. C. Xue, A. M. J. J. Bonvin, Protein-ligand pose and affinity prediction: Lessons from D3R Grand Challenge 3. *J. Comput. Aided Mol. Des.* **33**, 83–91 (2019). doi: 10.1007/s10822-018-0148-4; pmid: 30128928
69. S. Q. Zheng *et al*., MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017). doi: 10.1038/nmeth.4193; pmid: 28250466
70. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017). doi: 10.1038/nmeth.4169; pmid: 28165473
71. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016). doi: 10.1016/j.jsb.2015.11.003; pmid: 26592709
72. S. H. W. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012). doi: 10.1016/j.jsb.2012.09.006; pmid: 23000701
73. J. Zivanov *et al*., New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018). doi: 10.7554/eLife.42166; pmid: 30412051
74. A. Punjani, H. Zhang, D. J. Fleet, Non-uniform refinement: Adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020). doi: 10.1038/s41592-020-00990-8; pmid: 33257830
75. I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* **46**, 389–422 (2002). doi: 10.1023/A:1012487302797
76. C. Buchanan *et al*., Pathogen-sugar interactions revealed by universal saturation transfer analysis. Zenodo, doi: 10.5281/zenodo.6299883 (2022).doi: 10.5281/zenodo.6299883
77. P. I. Koukos, A. M. J. J. Bonvin, Modelling of the binding of various sialic acid-containing oligosaccharides to the NTD domain of the spike protein of SARS-CoV2. Zenodo, doi: 10.5281/zenodo.4271288 (2020).
78. C. J. Buchanan, A. J. Baldwin, Universal Saturation Transfer Analysis software. Zenodo, doi: 10.5281/zenodo.6303964 (2022).
79. N. Picchiotti *et al*., Post-Mendelian genetic model in COVID-19. *Cardiol. Cardiovasc. Med.* **5**, 673–694 (2021). doi: 10.26502/fccm.92920232
80. C. Fallerini *et al*., Common, low-frequency, rare, and ultra-rare coding variants contribute to COVID-19 severity. *Hum. Genet.* **141**, 147–173 (2022). doi: 10.1007/s00439-021-02397-7; pmid: 34889978

ACKNOWLEDGMENTS

We thank X. Chen (University of California, Davis) for providing the plasmid encoding Pd2,6ST; J. Mascola (NIH) for the plasmids encoding IAV NC99 (H1N1) HA variants; P. Supasa and G. Screaton (University of Oxford) for templates for alpha and beta spike; M. Kutuzov and O. Dushek for useful discussions and assistance with SPR experiments; A. Quigley and N. Gamage [Wellcome Trust Membrane Protein Lab, Diamond Light Source (20289/Z16/Z)] for help with thermal stability assays; C. Redfield for assistance with high-field Bruker instrumentation; and L. Baker for insights into the structural models. **Funding:** Upgrades of 600-MHz and 950-MHz spectrometers were funded by the Wellcome Trust (grant ref: 095872/Z/10/Z) and the Engineering and Physical Sciences Research Council (grant ref: EP/R029849/1), respectively, and by the University of Oxford Institutional Strategic Support Fund, the John Fell Fund, and the Edward Penley Abraham Cephalosporin Fund. We thank the Wellcome Trust for support (20289/Z/16/Z and 100209/Z/12/Z). I.B. is an Investigator of the Wellcome Trust (106115/Z/14/Z). Cryo-EM time was funded by a philanthropic gift to support COVID-19 research at the University of Oxford. The Chemistry theme at the Rosalind Franklin Institute is supported by the EPSRC (V011359/1 (P)). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 101002859). A.M.J.J.B. and P.I.K. acknowledge financial support from the European Union Horizon 2020 projects BioExcel (823830) and EOSC-Hub (777536) and the IMI-CARE project (101005077). This study is part of the GEN-COVID Multicenter Study (https://sites.google.com/dbm.unisi.it/gen-covid), the Italian multicenter study aimed at identifying the COVID-19 host genetic bases. Specimens were provided by the COVID-19 Biobank of Siena, which is part of the Genetic Biobank of Siena, member of BBMRI-IT, of Telethon Network of Genetic Biobanks (project GTB18001), of EuroBioBank, and of RDConnect. We thank the CINECA consortium for providing computational resources and the Network for Italian Genomes (NIG; www.nig.cineca.it) for its support. We thank private donors for the support provided to AR (Department of Medical Biotechnologies, University of Siena) for the COVID-19 host genetics research project (D.L n.18 of March 17, 2020). We also thank the COVID-19 Host Genetics Initiative (www.covid19hg.org/), MIUR project "Dipartimenti di Eccellenza 2018–2020" to the Department of Medical Biotechnologies University of Siena, Italy, and "Bando Ricerca COVID-19 Toscana" project to Azienda Ospedaliero–Universitaria Senese. We thank Intesa San Paolo for the 2020 charity fund dedicated to the project N.B/2020/0119 "Identificazione delle basi genetiche determinanti la variabilità clinica della risposta a COVID-19 nella popolazione italiana" and the Tuscany Region for funding within "Bando Ricerca COVID-19 Toscana" project supporting research at the Azienda Ospedaliero-Universitaria Senese. The Italian Ministry of University and Research for funding within the "Bando FISR 2020" in COVID-19 fo the project "Editing dell'RNA contro il Sars-CoV-2: hackerare il virus per identificare bersagli molecolari e attenuare l'infezione - HACKTHECOV" and the Istituto Buddista Italiano Soka Gakkai for funding the project "PAT-COVID: Host genetics and pathogenetic mechanisms of COVID-19" (ID n. 2020-2016_RIC_3). We thank EU project H2020-SC1-FA-DTS-2018-2020, entitled "International consortium for integrative genomics prediction (INTERVENE)" - grant agreement 101016775. **Author contributions:** B.G., C.J.B., M.J.D., T.D.W.C., and A.J.B. ran NMR experiments. A.J.B. and C.J.B. developed the underlying deconvolution algorithm, UnidecNMR. A.J.B. developed the modified Bloch-McConnell K_D analysis method. A.J.B. and C.J.B. performed the processing, uSTA applications, and surrounding analysis. C.J.B. and A.K. assigned key uSTA spectra. P.H. and A.L.B. purified spike protein and prepared samples for NMR analysis; P.H. and A.L.B. conducted thermal denaturation assays. A.K. synthesized 9-BPC-Neu5Ac, expressed and purified NmCSS and Pd2,6ST, assigned proton peaks in ^{1}H NMR spectra of sialosides, designed TOCSY NMR experiments, and performed the SPR experiments. A.M.G., A.K., and X.X. synthesized **5**, **6**, and **9**. AMG generated reagents for SPR studies. A.K., A.M.G., and A.J.B. analyzed SPR data. G.P. generated **7** and **8** and assigned their ^{1}H NMR spectra. J.H. cloned spike-BAP; P.W., M.D., L.S., and T.K.T. expressed spike protein. A.L. aided design of protein-ligand NMR experiments, prepared samples, and contributed to data analysis. A.R. is coordinating the GEN-COVID Consortium and supervised the genotype-phenotype correlation of glycosylation-associated genes. A.G. performed WES experiments. E.B. performed WES alignment and joint call, which was supervised by S.F. N.P. performed logistic regression analysis, which was supervised by M.G. M.B. and F.F. collected and supervised the collection of clinical data. C.F. and S.D. performed specific analysis on *B3GNT8* and *LGALS3BP* genes. L.P.D. and Q.J.S. performed HA expression and assisted with experimental design. A.M.J.J.B. and P.I.K. performed the structural modeling of the SARS-CoV-2 spike-sugar interactions. A.J.B. and A.M.J.J.B. scored these against the NMR data to obtain structural models for the complex. V.C.-A., C.J.B., A.J.B., and B.G.D. produced figures. J.H.N., Y.Y., and J.L. performed the EM experiments and analysis. I.B., C.S., and K.G. produced the protein used in EM. B.G.D., A.J.B., T.D.W.C., and J.H.N. supervised and designed experiments, analyzed data, and wrote the manuscript. All authors read and commented on the manuscript. **Competing interests:** The uSTA software is freely available to all academic users. In the event that the uSTA is licensed by a commercial party, C.J.B. and A.J.B. will then be afforded royalties in line with standard university practice. All authors declare that they have no other competing interests. **Data and materials availability:** The plasmid for the C5 nanobody is available on Addgene (plasmid #171925). Raw spectral data for Figs. 1 to 5 and the data used to plot Fig. 4, A to C have been deposited (*76*). Associated spectral data are also shown in table S7. HADDOCK datasets and results have been deposited (*77*). All associated clinical data have been supplied in tables S1 to S6, conducted under trial (NCT04549831, www.clinicaltrial.org). The uSTA software is free for academic use and has been deposited (*78*). The coordinates and EM maps have been deposited with the RCSB: Spike_native_C1 (EMDB-14155), Spike_native_C3 (EMDB-14153) (PDB 7QUS). Spike_ sialic acid _C1 (EMDB-14154) and Spike_ sialic acid_C3 (EMDB-14152) (PDB 7QUR).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm3125
Supplementary Methods
Supplementary Text
Figs. S1 to S23
Tables S1 to S12
References (*81*–*102*)

Submitted 16 September 2021; accepted 16 June 2022
Published online 23 June 2022
10.1126/science.abm3125

RESEARCH ARTICLE SUMMARY

PLANT SCIENCE

A transcriptional regulator that boosts grain yields and shortens the growth duration of rice

Shaobo Wei†, Xia Li†, Zefu Lu, Hui Zhang, Xiangyuan Ye, Yujie Zhou, Jing Li, Yanyan Yan, Hongcui Pei, Fengying Duan, Danying Wang, Song Chen, Peng Wang, Chao Zhang, Lianguang Shang, Yue Zhou, Peng Yan, Ming Zhao, Jirong Huang, Ralph Bock, Qian Qian, Wenbin Zhou*

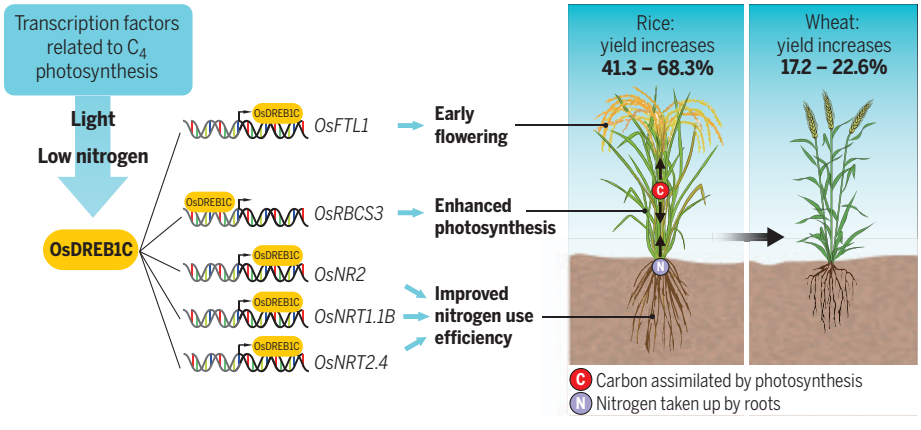
INTRODUCTION: Rapid population growth, rising meat consumption, and the expanding use of crops for nonfood and nonfeed purposes increase the pressure on global food production. At the same time, the excessive use of nitrogen fertilizer to enhance agricultural production poses serious threats to both human health and the environment. To achieve the required yield increases and make agriculture more sustainable, intensified breeding and genetic engineering efforts are needed to obtain new crop varieties with higher photosynthetic capacity and improved nitrogen use efficiency (NUE). However, progress has been slow, largely due to the limited knowledge about regulator genes that potentially can coordinately optimize carbon assimilation and nitrogen utilization.

RATIONALE: Transcription factors control diverse biological processes by binding to the promoters (or intragenic regions) of target genes, and a number of transcription factors have been identified that control carbon fixation and nitrogen assimilation. A previous comparative analysis of maize and rice leaf transcriptomes and metabolomes revealed a set of 118 candidate transcription factors that may act as regulators of C_4 photosynthesis. We screened these transcription factors for their responsiveness to light and nitrogen supply in rice, and found that the gene *Dehydration-Responsive Element-Binding Protein 1C* (*OsDREB1C*), a member of the APETALA2/ethylene-responsive element binding factor (AP2/ERF) family, exhibits properties expected of a regulator that can simultaneously modulate photosynthesis and nitrogen utilization.

RESULTS: *OsDREB1C* expression is induced in rice by both light and low-nitrogen status. We generated overexpression lines (*OsDREB1C*-OE) and knockout mutants (*OsDREB1C*-KO) in rice, and conducted field trials in northern, southeastern, and southern China from 2018 to 2021. *OsDREB1C*-OE plants exhibited 41.3 to 68.3% higher yield than wild-type (WT) plants due to increased grain number per panicle, elevated grain weight, and enhanced harvest index. We observed that light-induced growth promotion of *OsDREB1C*-OE plants was accompanied by enhanced photosynthetic capacity and concomitant increases in photosynthetic assimilates. In addition, ^{15}N feeding experiments and field studies with different nitrogen fertilization regimes revealed that NUE was improved in *OsDREB1C*-OE plants due to elevated nitrogen uptake and transport activity. Moreover, *OsDREB1C* overexpression led to more efficient carbon and nitrogen allocation from source to sink, thus boosting grain yield, particularly under low-nitrogen conditions. Additionally, the *OsDREB1C*-OE plants flowered 13 to 19 days earlier and accumulated higher biomass at the heading stage than WT plants under long-day conditions.

OsDREB1C is localized in the nucleus and the cytosol and functions as a transcriptional activator that directly binds to *cis* elements in the DNA, including dehydration-responsive element (DRE)/C repeat (CRT), GCC, and G boxes. Chromatin immunoprecipitation sequencing (ChIP-seq) and transcriptomic analyses identified a total of 9735 putative OsDREB1C-binding sites at the genome-wide level. We discovered that five genes targeted by OsDREB1C [*ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit 3* (*OsRBCS3*), *nitrate reductase 2* (*OsNR2*), *nitrate transporter 2.4* (*OsNRT2.4*), *nitrate transporter 1.1B* (*OsNRT1.1B*), and *flowering locus T-like 1* (*OsFTL1*)] are closely associated with photosynthesis, nitrogen utilization, and flowering, the key traits altered by *OsDREB1C* overexpression. ChIP-quantitative polymerase chain reaction (ChIP-qPCR) and DNA affinity purification sequencing (DAP-seq) assays confirmed that OsDREB1C activates the transcription of these genes by binding to the promoter of *OsRBCS3* and to exons of *OsNR2*, *OsNRT2.4*, *OsNRT1.1B*, and *OsFTL1*. By showing that biomass and yield increases can also be achieved by OsDREB1C overexpression in wheat and *Arabidopsis*, we have demonstrated that the mode of action and the biological function of the transcription factor are evolutionarily conserved.

CONCLUSION: Overexpression of *OsDREB1C* not only boosts grain yields but also confers higher NUE and early flowering. Our work demonstrates that by genetically modulating the expression of a single transcriptional regulator gene, substantial yield increases can be achieved while the growth duration of the crop is shortened. The existing natural allelic variation in *OsDREB1C*, the highly conserved function of the transcription factor in seed plants, and the ease with which its expression can be altered by genetic engineering suggest that this gene could be the target of future crop improvement strategies toward more efficient and more sustainable food production. ■



OsDREB1C coordinates yield and growth duration. OsDREB1C was identified by its responsiveness to light and low nitrogen in a screen of 118 transcription factors related to C_4 photosynthesis. Transcriptional activation of multiple downstream target genes by OsDREB1C confers enhanced photosynthesis, improved nitrogen utilization, and early flowering. Together, the activated genes cause substantial yield increases in rice and wheat.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: zhouwenbin@caas.cn
†These authors contributed equally to this work.
Cite this article as S. Wei *et al.*, *Science* 377, eabi8455 (2022). DOI: 10.1126/science.abi8455

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abi8455

RESEARCH ARTICLE

PLANT SCIENCE

A transcriptional regulator that boosts grain yields and shortens the growth duration of rice

Shaobo Wei[1]†, Xia Li[1]†, Zefu Lu[1], Hui Zhang[2], Xiangyuan Ye[1], Yujie Zhou[1], Jing Li[1], Yanyan Yan[1], Hongcui Pei[1], Fengying Duan[1], Danying Wang[3], Song Chen[3], Peng Wang[4], Chao Zhang[5], Lianguang Shang[5], Yue Zhou[6], Peng Yan[6], Ming Zhao[1], Jirong Huang[2], Ralph Bock[7], Qian Qian[1,3], Wenbin Zhou[1]*

Complex biological processes such as plant growth and development are often under the control of transcription factors that regulate the expression of large sets of genes and activate subordinate transcription factors in a cascade-like fashion. Here, by screening candidate photosynthesis-related transcription factors in rice, we identified a DREB (Dehydration Responsive Element Binding) family member, OsDREB1C, in which expression is induced by both light and low nitrogen status. We show that OsDREB1C drives functionally diverse transcriptional programs determining photosynthetic capacity, nitrogen utilization, and flowering time. Field trials with *OsDREB1C*-overexpressing rice revealed yield increases of 41.3 to 68.3% and, in addition, shortened growth duration, improved nitrogen use efficiency, and promoted efficient resource allocation, thus providing a strategy toward achieving much-needed increases in agricultural productivity.

Globally, >800 million people are suffering from hunger and food insecurity (*1*). By 2050, crop production needs to be increased by 50 to 70% to feed nearly 10 billion people despite the reduced availability of arable land on the planet (*2*, *3*). Meeting this challenge will likely require the development of new breeding and genetic engineering strategies that optimize photosynthetic capacity as well as water and nutrient use efficiency (*4*). Growth and crop yield depend on carbon and nitrogen assimilation and photosynthate translocation from vegetative source organs to sink tissues (*5*, *6*). For example, nitrogen uptake and transport must be coordinated with carbon fixation and the production of carbohydrates by photosynthesis. Therefore, research efforts have been devoted to identifying transcriptional regulators that control the coordination between carbon assimilation and nitrogen utilization (*7*, *8*).

As the regulators of biological processes, transcription factors control plant metabolism, growth, and development by binding to the promoters (or intragenic regions) of target genes (*9*, *10*). An example of a transcription factor in plant architecture is TB1 (Teosinte Branched 1) of maize, which limits branch outgrowth and initiates the formation of female inflorescences (*11*, *12*). The transcription factor IPA1 (Ideal Plant Architecture 1) promotes rice yield by reducing unproductive tillers and increasing grain number per panicle. Elevated IPA1 levels also enhance pathogen immunity (*13*, *14*). Another transcription factor, HYR (HIGHER YIELD RICE), enhances the expression of photosynthesis genes and can increase rice yield under multiple stress conditions (*15*). The rice transcription factor GRF4 (GROWTH-REGULATING FACTOR4) coordinates nitrogen assimilation, carbon fixation, and growth (*7*). Often, binding motifs and functions of transcription factors are conserved in monocot and dicot species (*7*, *16*).

Previous work has been directed at identifying key transcription factors that regulate photosynthesis, and nitrogen and carbon metabolism, using comparative analysis of maize and rice leaf transcriptomes and metabolomes (*17*). A set of 118 transcription factors were considered as candidate regulators of photosynthesis and, especially, of favorable properties related to C_4 photosynthesis (*17*). Here, we screened these transcription factors for their responsiveness to light and nitrogen supply in rice. We report the identification of a transcription factor from the DREB (Dehydration Responsive Element Binding) family, OsDREB1C, that modulates both photosynthesis and nitrogen utilization. Overexpression of *OsDREB1C* not only increases rice yields, but also confers early grain maturation because of higher rates of photosynthesis, improved nitrogen utilization, and early flowering. The versatile functions of OsDREB1C are likely conferred by the transcription factor acting near the top of the hierarchy and coordinately targeting multiple genes and pathways. Recognition of the conserved dehydration-responsive element (DRE)/C repeat (CRT) motif present in these loci facilitates fine-tuning of the intricate networks of carbon assimilation, nitrogen utilization, resource allocation, and induction of flowering. Our results uncover OsDREB1C as a transcriptional regulator that promotes the expression of key genes in carbon and nitrogen metabolism and controls flowering pathways in crops, thus providing a target for future crop improvement strategies.

RESULTS

OsDREB1C boosts grain yield and harvest index

To investigate how the coordination of photosynthesis and nitrogen utilization affects grain yield in cereals, we conducted RNA-sequencing (RNA-seq) analyses in rice plants grown under low- versus high-nitrogen conditions (*18*). We examined the differential expression of the subset of transcription factors associated with photosynthesis gene expression in maize (*17*). We detected nitrogen-regulated expression for 13 of these transcription factors, with five genes showing a greater than fourfold induction under low-nitrogen conditions (Fig. 1A). Investigation of light-regulated mRNA accumulation showed that one of the five genes, *Os06g0127100* (encoding a DREB-type transcription factor previously shown to be inducible by abiotic stress), displayed a diurnal rather than a circadian expression profile, with expression increasing with the duration of light exposure (Fig. 1B and fig. S1).

To facilitate the functional analysis of this apparently nitrogen-regulated and light-induced transcription factor, we generated a series of *OsDREB1C*-overexpressing lines (*OsDREB1C*-OE) and *OsDREB1C*-knockout mutants (*OsDREB1C*-KO) in the *Oryza sativa* cv. Nipponbare genetic background (fig. S2). Field tests of these plant lines in Beijing in 2018 revealed that *OsDREB1C* overexpression led to increases in grain yield per plant of 45.1 to 67.6% and in yield per plot of 41.3 to 68.3% compared with wild-type (WT) plants (Fig. 1C). Conversely, *OsDREB1C* KO resulted in yield decreases (from 16.1 to 29.1% in yield per plant and 13.8 to 27.8% in yield per plot) compared with the WT (Fig. 1, D and E, and table S1). A detailed phenotypic analysis showed that the higher yield of the *OsDREB1C*-OE lines was mainly attributable to an enhanced grain number per panicle and an increased 1000-grain weight (Fig. 1F and fig. S3C), traits apparently resulting from increased

[1]Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China. [2]Shanghai Key Laboratory of Plant Molecular Sciences, College of Life Sciences, Shanghai Normal University, Shanghai 200234, China. [3]State Key Laboratory of Rice Biology, China National Rice Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou 310006, China. [4]CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200032, China. [5]Lingnan Laboratory of Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518124, China. [6]State Key Laboratory of Protein and Plant Gene Research, School of Advanced Agricultural Sciences, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China. [7]Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg, 14476 Potsdam-Golm, Germany.
*Corresponding author. Email: zhouwenbin@caas.cn
†These authors contributed equally to this work.

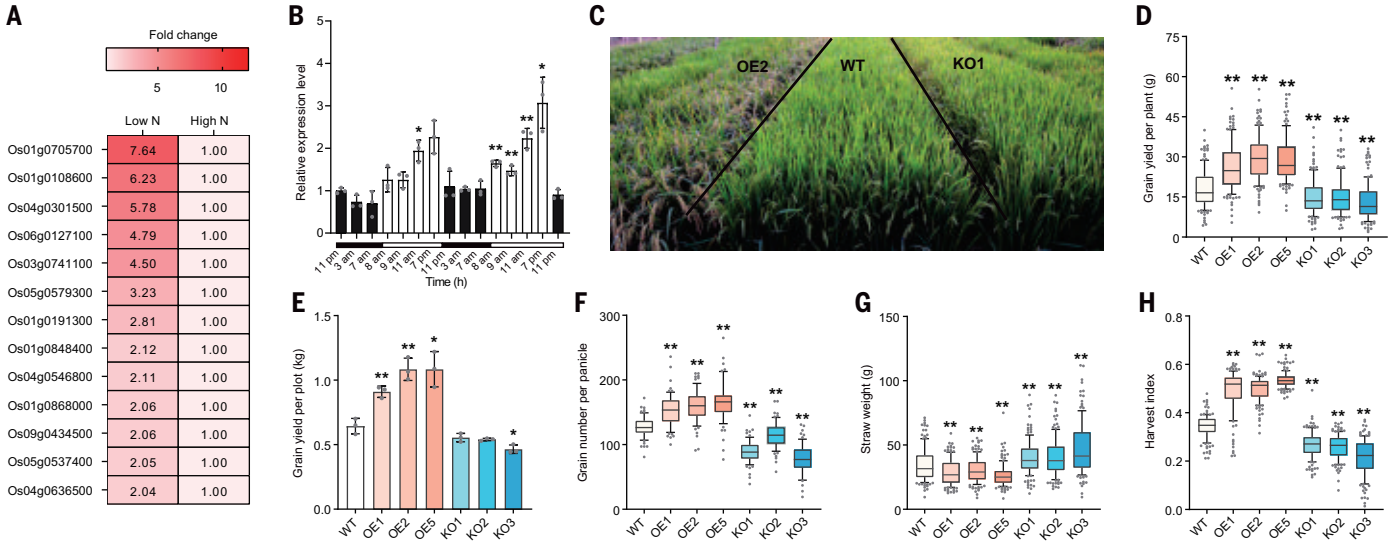


Fig. 1. *OsDREB1C* overexpression in transgenic plants boosts grain yield. (**A**) List of the top 13 genes up-regulated in response to nitrogen deprivation (adjusted $P < 0.05$). The genes represent the overlap of previously reported RNA-seq datasets (*17*) and an expression analysis of a subset of 118 rice transcription factors (*16*), and were sorted by the fold change in low versus normal nitrogen supply. The color scale represents the $\log_2$-fold change of the FPKM (fragments per kilobase of transcript per million mapped reads) ratio under low- versus high-nitrogen conditions, with the FPKM value of each gene under high-nitrogen conditions set to 1.00. (**B**) qRT-PCR analysis of *Os06 g0127100* expression in 10-day-old *O. sativa* cv. Nipponbare seedlings grown in soil in a growth chamber under long-day photoperiod (16 hours light/8 hours dark, 28°C). The white bar below the *x*-axis indicates the light period, and the black bar indicates the dark period. Data are presented as means ± SD ($n = 3$ biological replicates). $^*P < 0.05$, $^{**}P < 0.01$ compared with the first time point (11:00 p.m.), Student's *t* test. (**C**) Phenotypes of WT and transgenic rice plants grown in Beijing in 2018. (**D** to **H**) Yield-related parameters including grain yield per plant (D), grain yield per plot (E), grain number per panicle (F), straw weight (G), and harvest index (H). The data were obtained from the field experiment shown in (C). Box plots in (D) and (F) to (H) show median (horizontal lines) and 10th to 90th percentiles, and outliers are plotted as dots ($n = 138$ biological replicates). Data in (E) are presented as means ± SD ($n = 3$ plots, 44 plants within a plot). $^*P < 0.05$, $^{**}P < 0.01$ compared with WT, Student's *t* tests.

secondary branch number and grain length, width, thickness, and density (fig. S3, A and B and D to K, respectively). The *OsDREB1C*-OE plants exhibited higher grain yield but reduced straw weight compared with WT plants (Fig. 1G), thus leading to an increased harvest index (the ratio of grain yield to aboveground biomass; Fig. 1H) and raising the possibility that OsDREB1C controls resource allocation between vegetative and reproductive tissues. The harvest index of *OsDREB1C*-OE plants was increased by 40.3 to 55.7%, whereas it was decreased by 22.4 to 33.7% in *OsDREB1C*-KO plants (table S1). In addition, key grain quality traits were enhanced in *OsDREB1C*-OE plants, suggesting that yield improvement does not entail a quality penalty (table S2).

To assess the stability of the yield enhancement conferred by OsDREB1C, we conducted field trials over several years and at three different sites that represent very different environmental conditions (tables S1 and S3 to S5). Data for the Beijing field trial in 2019 showed an even larger increase in grain yield for *OsDREB1C*-OE plants than in 2018 and similar yield reductions in *OsDREB1C*-KO plants (table S3). Having tested in temperate (Beijing; tables S1 and S3 and fig. S3M), tropical (Hainan Province; table S4 and fig. S3N), and subtropical (Zhejiang Province; table S5) locations, we noted that the most pronounced yield increases were detected in the long-day photoperiod and temperate climate conditions of Beijing. Nevertheless, there was still a strong yield improvement for *OsDREB1C*-OE plants in the short-day and tropical conditions of Hainan, with yield increases in the range of 7.8 to 16% yield per plant and 12.0 to 37.0% yield per plot (table S4 and fig. S3N). Also, the harvest index of *OsDREB1C*-OE plants (~0.62) was higher than in WT (~0.54) and *OsDREB1C*-KO (~0.32) plants in Hainan (table S4).

OsDREB1C improves photosynthetic capacity and nitrogen utilization

To further explore the molecular basis of the yield enhancement conferred by *OsDREB1C* overexpression, a series of physiological measurements were conducted with hydroponically grown WT and transgenic plants. In these experiments, we observed faster growth of *OsDREB1C*-OE plants already at the seedling stage, whereas the growth of *OsDREB1C*-KO seedlings was retarded compared with WT plants (fig. S4, A and B). In addition, we noticed that *OsDREB1C*-OE plants displayed longer roots, probably related to their increased auxin content (fig. S5). We detected no growth differences among the WT, overexpression, and KO plants when the seedlings were cultivated in the dark (figs. S4, C to F, and S6), although the *OsDREB1C*-OE lines showed taller shoots during the first 10 days, likely because of the larger grain size, which provides more reserves for initial growth (fig. S6). Overall, these data suggested a light-induced mechanism of growth improvement. We next investigated whether photosynthetic capacity is improved by the overexpression of *OsDREB1C*. Leaves of *OsDREB1C*-OE plants contained higher levels of photosynthetic pigments (chlorophylls and carotenoids) compared with WT plants, whereas pigment levels were reduced in *OsDREB1C*-KO plants (fig. S7A). Analysis of leaf mesophyll cells revealed that both chloroplast number and size were increased in *OsDREB1C*-OE plants (fig. S7, B and C). Biochemical analysis of photosynthetic protein complexes by blue-native polyacrylamide gel electrophoresis revealed elevated levels of photosystem I (PSI) and PSII dimers, PSII-CP43 monomers, and light-harvesting complex II (LHCII) trimers in *OsDREB1C*-OE plants (fig. S7D). Immunoblotting confirmed an increased abundance of PSI, PSII, cytochrome b_6/f, ATP synthase, and LHC proteins in leaves of the overexpression lines (fig. S7E). *OsDREB1C* overexpression also led to enhanced amounts of the large and small subunits of ribulose bisphosphate carboxylase/oxygenase (RbcL and RbcS, respectively) and ribulose-1,5-bisphosphate

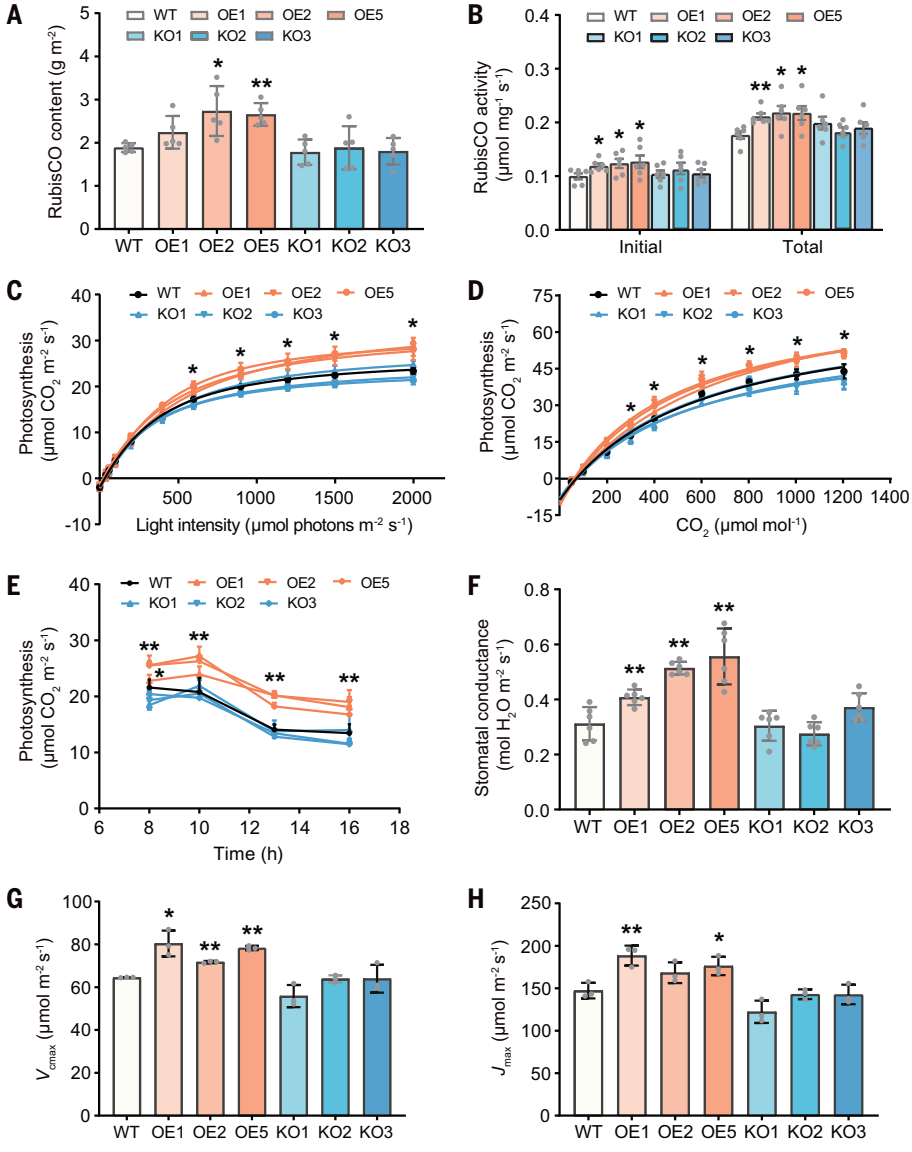


Fig. 2. *OsDREB1C* overexpression promotes photosynthetic capacity. (**A** and **B**) RuBisCO content (A) and RuBisCO activity (B) in the fourth leaf of 3-week-old rice seedlings grown hydroponically. Data are presented as means ± SD (n = 6 biological replicates). (**C**) Light-response curve of net photosynthesis fitted by the Farquhar–von Caemmerer–Berry (FvCB) model and generated at 30°C and 400 ppm CO_2 concentration in the field in Hainan in 2019. Data are presented as means ± SD (n = 3 biological replicates). *P < 0.05 for OE1, OE2, and OE5 compared with WT, Student's t test. (**D**) CO_2-response curve of net photosynthesis (A-C_i curve) fitted by the FvCB model and generated at 1200 μmol m^{-2} s^{-1} photosynthetic photon flux density and 30°C. Data are presented as means ± SD (n = 3 biological replicates). *P < 0.05 for OE1, OE2, and OE5 compared with WT, Student's t test. (**E**) Diurnal changes in the photosynthesis rate of flag leaves measured with a LICOR-6400 XT instrument at the heading stage (from 8:00 a.m. to 4:00 p.m.) in the field in Hainan. **P < 0.01 for OE1, OE2, and OE5 compared with WT, *P < 0.05 for OE1 at 8:00 a.m., all Student's t test. (**F**) Stomatal conductance of flag leaves at the heading stage in Hainan. Measurements were performed with a LICOR-6400 XT instrument at 30°C in ambient air between 10:00 a.m. and 11:00 a.m. Data in (E) and (F) are presented as presented as means ± SD (n = 6 biological replicates). (**G** and **H**) Maximum rates of carboxylation (V_{cmax}) (G) and electron transport (J_{max}) (H) in WT, *OsDREB1C*-OE, and *OsDREB1C*-KO plants grown in the field in Hainan. Values were generated from the A-C_i curve and fitted by the FvCB model. Data are presented as means ± SD (n = 3 biological replicates). All measurements in (C) to (H) were performed with flag leaves of field-grown rice plants at the heading stage. *P < 0.05, **P < 0.01 compared with WT, Student's t test.

carboxylase-oxygenase (RuBisCO) activase (fig. S7E). Accordingly, both RuBisCO content and activity were increased in *OsDREB1C*-OE plants (Fig. 2, A and B).

Next, we evaluated the role of *OsDREB1C* in regulating photosynthetic capacity by investigating rice plants grown in paddy fields. Consistent with the above-described results, key photosynthetic parameters, including diurnal changes, light-response curves, and CO_2-response curves of net photosynthesis, were improved in *OsDREB1C*-OE plants and compromised in *OsDREB1C*-KO plants (Fig. 2, C to E). In addition, *OsDREB1C*-OE plants displayed higher stomatal conductance while maintaining a similar intercellular CO_2 concentration (Fig. 2F and fig. S8), suggesting that the higher carboxylation rates supported by higher RuBisCO content and activity prevent the buildup of higher intercellular CO_2 levels despite the opened stomata. This is in agreement with the higher maximum rate of RuBisCO carboxylation and the higher maximum rate of electron transport derived from modeling of A–C_i curves (Fig. 2, G and H). Analysis of key products of photosynthetic metabolism showed that *OsDREB1C*-OE leaves accumulated higher amounts of starch, sucrose, and fructose, thus potentially explaining the improved grain filling (fig. S9). Together, these results suggest that the observed growth and yield increases in *OsDREB1C*-OE rice plants result, at least in part, from enhancement of photosynthetic capacity.

To investigate whether *OsDREB1C* also influences nitrogen utilization, we conducted a ^{15}N feeding experiment and monitored nitrogen uptake and transport activity in hydroponically grown seedlings. After 3 hours of incubation in a ^{15}N-nitrate solution, *OsDREB1C*-OE seedlings had higher ^{15}N contents in shoots and roots compared with WT and *OsDREB1C*-KO plants (Fig. 3, A and B). Additionally, *OsDREB1C* overexpression increased both nitrogen uptake by roots and transport activity from roots to shoots (Fig. 3, C and D), thus resulting in elevated nitrogen content, nitrogen use efficiency (NUE), and protein abundance in leaves of field-grown plants (Fig. 3, E and F, and fig. S10). Consistent with these findings, mature field-grown *OsDREB1C*-OE plants had increased total nitrogen contents in above-ground organs (Fig. 3G and fig. S11A), along with improved photosynthetic NUE in a range of nitrogen supply conditions and a higher intrinsic water use efficiency upon low-level nitrogen application (fig. S12), indicating higher efficiency of carbon gain at lower nitrogen cost. Analysis of carbon and nitrogen distribution showed that the *OsDREB1C*-OE plants accumulated more carbon and nitrogen in the grains, but less in

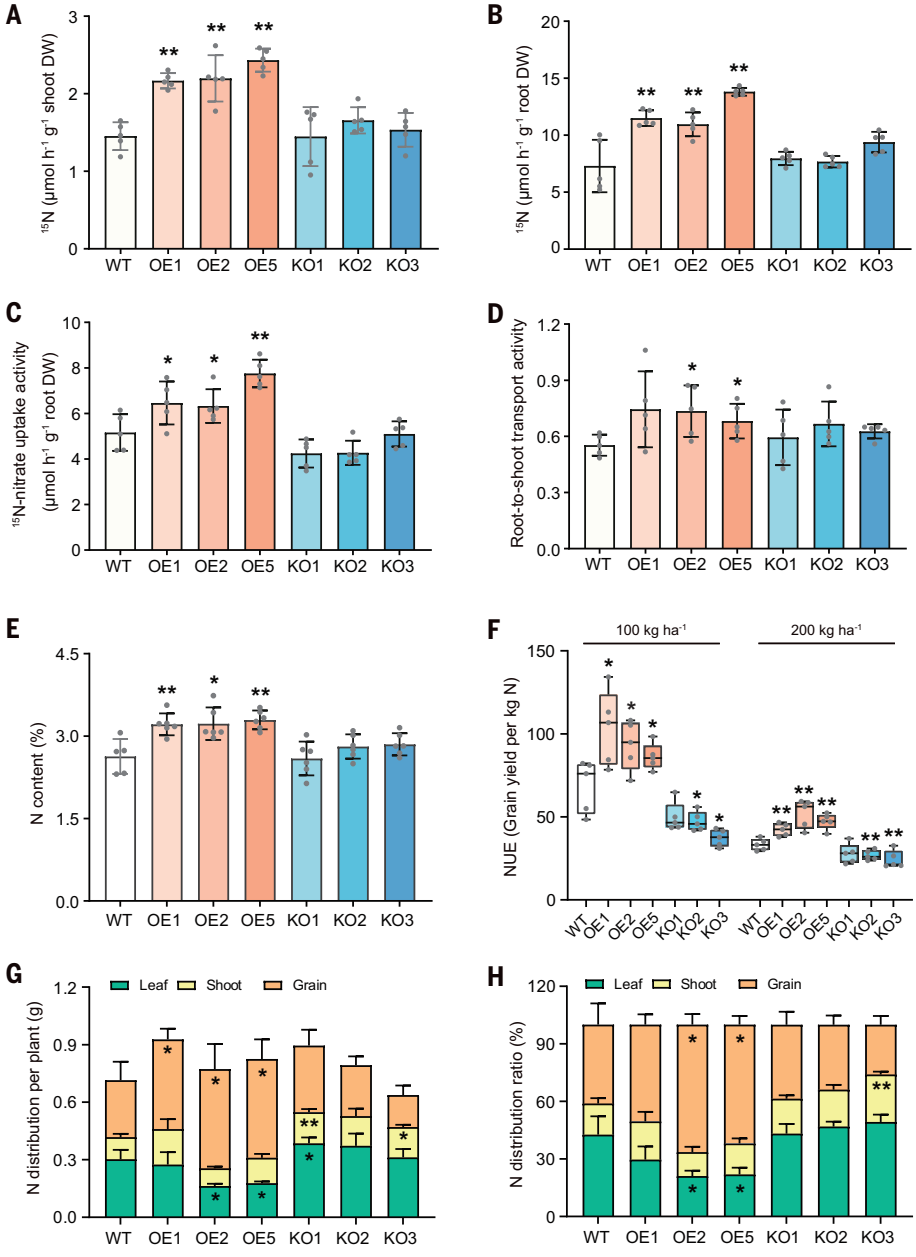


Fig. 3. *OsDREB1C* overexpression increases nitrogen uptake and transport. (**A** and **B**) ^{15}N content in shoots (A) and roots (B) of 3-week-old WT, *OsDREB1C*-OE, and *OsDREB1C*-KO seedlings incubated with 0.5 mM $K^{15}NO_3$ for 3 hours. (**C**) ^{15}N-nitrate uptake activity of roots. (**D**) ^{15}N transport activity from roots to shoots. Data in (A) to (D) are presented as means ± SD (n = 5 biological replicates). (**E**) Nitrogen content of flag leaves of WT, *OsDREB1C*-OE, and *OsDREB1C*-KO plants at the heading stage grown in the field in Beijing in 2021. Data are presented as means ± SD (n > 5 biological replicates). (**F**) NUE of WT, *OsDREB1C*-OE, and *OsDREB1C*-KO plants grown with 100 or 200 kg ha^{-1} nitrogen supply in the field in Beijing in 2021. Box plot shows median (line) and individual values (black dots) (n > 5 biological replicates). *P < 0.05, **P < 0.01 compared with WT, Student's t test. (**G** and **H**) Nitrogen distribution (G) and nitrogen distribution ratio (H) in seeds, straw, and leaves of mature plants grown in the field in Beijing in 2019. Data are presented as means ± SD (n = 4 biological replicates). *P < 0.05, **P < 0.01 compared with WT, Student's t test.

their mature leaves, without substantial alterations in the carbon-to-nitrogen ratio (Fig. 3H and fig. S11). These findings indicate more efficient resource allocation from source (leaves) to sink (grains) in the overexpression plants compared with WT plants (fig. S11), leading to a further boost in grain yield, especially under conditions of low nitrogen supply (fig. S13). In conclusion, *OsDREB1C* appears to (i) stimulate nitrogen uptake by roots, (ii) enhance nitrogen transport to aerial organs, and (iii) promote resource allocation from leaves and shoots to grains.

OsDREB1C promotes flowering and shortens growth duration

OsDREB1C-OE plants also exhibited a pronounced early flowering phenotype in our field trials (Fig. 4, A and B). Early flowering was accompanying with higher biomass accumulation at the heading stage (Fig. 4C). *OsDREB1C*-OE plants flowered 13 to 19 days earlier than the WT, whereas *OsDREB1C*-KO plants flowered 3 to 7 days later under the long-day conditions in the field in Beijing (Fig. 4B). Overexpression of *OsDREB1C* also had an effect on leaf maturation and duration of the growth period (Fig. 4D). However, these differences were less evident under the short photoperiod in Hainan (table S4).

To further explore the influence of OsDREB1C on flowering, the expression of key genes involved in the induction of flowering was analyzed, including the two *FT*-like genes, *Hd3a* (*OsFTL2*) and *RFT1* (*OsFTL3*) (*19*, *20*), and the downstream transcription factor gene *OsMADS14*. Expression of all three genes was up-regulated in *OsDREB1C*-OE plants and down-regulated in *OsDREB1C*-KO plants (Fig. 4, E to G). By contrast, expression of *Hd1* and *Ehd1*, two genes upstream of florigen, was not affected in *OsDREB1C*-OE and *OsDREB1C*-KO plants (Fig. 4, H and I), indicating that OsDREB1C affects flowering time by controlling a specific subset of flowering regulators.

OsDREB1C enhances yield in an elite cultivar

To further test whether *OsDREB1C* overexpression can increase the yield of elite rice varieties, we transformed the *p35S::OsDREB1C* construct into Xiushui 134 (XS134), a high-yielding elite temperate *japonica* cultivar that is widely cultivated in southern China. Transgenic XS134 rice plants (*OsDREB1C*-XSOE) exhibited increased height, longer panicles, higher grain numbers per panicle, and higher grain yields over 2 consecutive years, similar to those seen in *OsDREB1C*-OE plants in the Nipponbare background (Fig. 5, A to G, and figs. S14 and S15). The grain yield per plot was increased by 10.3 to 12.7% in 2020 and 30.1 to 41.6% in 2021 in Hangzhou (fig. S14G and Fig. 5H), accompanied by an increased harvest index of up to 10.5 and 15.7%, respectively (fig. S14H and Fig. 5I). Moreover, *OsDREB1C*-XSOE lines flowered 2 days earlier than the WT in Hangzhou (fig. S14I). Accordingly, *OsDREB1C*-XSOE lines displayed 26.2 to 42.4% higher grain yields per plant in Hainan (fig. S15), whereas no difference in flowering time was observed because of the short-day growth conditions.

Next, we analyzed the *OsDREB1C* sequences in 709 rice accessions, including 299 *indica*, 355 temperate *japonica*, 14 tropical *japonica*, and 41 intermediate varieties (*21*). Three distinct

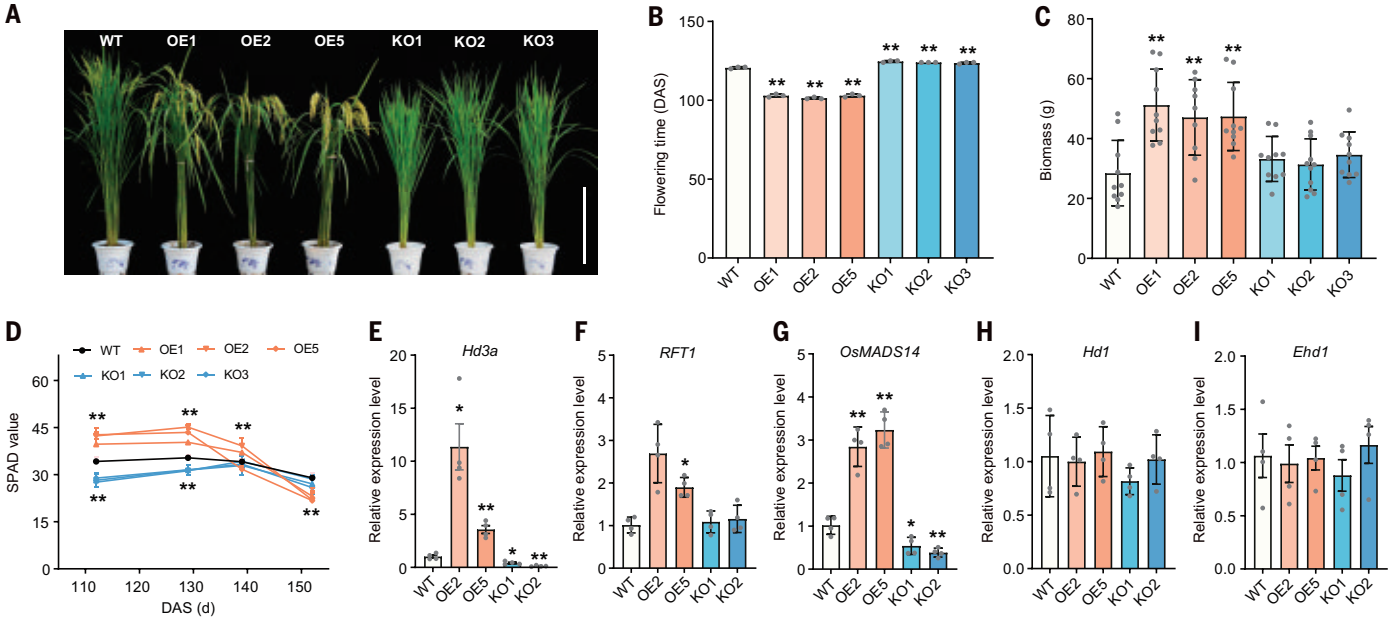


Fig. 4. *OsDREB1C* overexpression leads to early flowering and shortens the overall growth period. (**A**) Growth of WT, *OsDREB1C*-OE, and *OsDREB1C*-KO plants in natural long-day conditions in Beijing in 2019. Scale bar, 50 cm. (**B**) Flowering time of field-grown plants. DAS, days after sowing. (**C**) Biomass of rice plants at the heading stage of *OsDREB1C*-OE plants (105 DAS) grown in the field in Beijing in 2021. (**D**) Soil plant analysis development (SPAD) value of flag leaves at different development stages (112, 129, 139, and 152 DAS) of plants grown in the field in Beijing in 2019. $^{**}P < 0.01$ for OE1, OE2, and OE5 at four time points and for KO1, KO2, and KO3 at 112 and 129 DAS compared with WT, all Student's *t* test. (**E** to **I**) Relative gene expression levels of the flowering regulators *Hd3a* (E), *RFT1* (F), *OsMADS14* (G), *Hd1* (H), and *Ehd1* (I) in WT, *OsDREB1C*-OE, and *OsDREB1C*-KO plants. RNAs were extracted from leaves of field-grown plants before heading (~90 DAS) in the Beijing field in 2019. Data in (B) to (I) are presented as means ± SD [n = 3 biological replicates, except for n =10 for (C)]. $^{*}P < 0.05$, $^{**}P < 0.01$ compared with WT, Student's *t* test.

haplotypes were identified (Hap. 1 to Hap. 3) on the basis of nucleotide polymorphisms, most of which reside in the promoter regions (fig. S16).

Identification of target genes of OsDREB1C

We next wanted to characterize the molecular functions of OsDREB1C in more detail. Multiple amino acid sequence alignment revealed a conserved AP2 domain among all OsDREB1C homologs (figs. S17 and S18). Expression analysis showed that *OsDREB1C* was expressed ubiquitously in all rice tissues examined (root, stem, leaf, and panicle), but particularly strongly in the root (fig. S19A). During the growth period, *OsDREB1C* transcript levels peaked at the tillering stage (fig. S19B). Transient expression assays in rice protoplasts revealed that OsDREB1C-GFP and YFP-OsDREB1C fusion proteins mainly localize to the nucleus, but a substantially weaker signal in the cytoplasm was also discernable (fig. S19C). Very similar patterns of OsDREB1C-GFP subcellular localization were observed in *Arabidopsis* and *Nicotiana benthamiana* (fig. S19, D and E).

Sequence analysis with PlantPan3.0 suggested that the OsDREB1C protein may bind to the DRE/CRT (GCCGAC) motif that had been identified as a core *cis*-acting element regulating gene expression in response to drought, salt, and cold stresses (*22*). Yeast one-hybrid assays verified the direct binding of OsDREB1C to DRE/CRT (GCCGAC) as well as the GCC box (GCCGCC) and G box (CACGTG) *cis* elements in vitro (fig. S20A). Measurement of the transcription-stimulating activity in rice protoplasts demonstrated that OsDREB1C was able to activate transcription of the GUS reporter gene (fig. S20, B and C). Taken together, these data suggest that OsDREB1C functions as a transcriptional activator.

To identify genome-wide binding sites of OsDREB1C in vivo, we performed chromatin immunoprecipitation sequencing (ChIP-seq) experiments with rice protoplasts transiently expressing the OsDREB1C-GFP fusion protein. These analyses identified a total of 9735 putative OsDREB1C-binding sites, of which 68% localized to genic regions and 32% to intergenic regions (Fig. 6A). The core motif found to be enriched in the OsDREB1C-binding regions was DRE/CRT (GCCGAC) (Fig. 6B). We next analyzed the differentially expressed genes (DEGs) between *OsDREB1C*-OE and WT plants (as determined by RNA-seq), and extracted the DEGs associated with OsDREB1C-binding peaks. In this way, 345 up-regulated genes were identified as putative OsDREB1C targets (Fig. 6C). Gene ontology (GO) enrichment analysis was then conducted to associate biological processes with those DEGs. Transmembrane transport–related genes were found to be most strongly enriched and included the two nitrogen transporter genes, *OsNRT2.4* and *OsNRT1.1B* (Fig. 6D). Other genes with functions in the nitrogen metabolic process were also present in the DEG set, including the nitrate reductase gene *OsNR2*. When searching for flowering-related DEGs that could potentially explain the pronounced early-flowering phenotype, the gene *OsFTL1* (*FT-Like 1*) was found. *OsFTL1* is a homolog of the *Arabidopsis FT* gene that plays a central role in integrating signals from the different flowering pathways (*23*, *24*). Moreover, a key photosynthesis-related gene, *OsRBCS3*, encoding the RuBisCO small subunit and known to be transcriptionally activated by light (*25*), was also up-regulated in *OsDREB1C*-OE plants.

OsDREB1C directly activates key pathway genes

To verify whether these five candidate genes (*OsRBCS3*, *OsNR2*, *OsNRT2.4*, *OsNRT1.1B*, and *OsFTL1*) were targets of OsDREB1C, we performed ChIP–quantitative polymerase chain reaction (qPCR) experiments using transgenic plants expressing an OsDREB1C-GFP fusion protein and DNA affinity purification sequencing (DAP-seq) assays in vitro. The results revealed that OsDREB1C directly binds to the promoter of *OsRBCS3* and to exons of *OsNR2*, *OsNRT2.4*, *OsNRT1.1B*, and *OsFTL1* (Fig. 6E and fig. S21A). Electrophoretic mobility shift

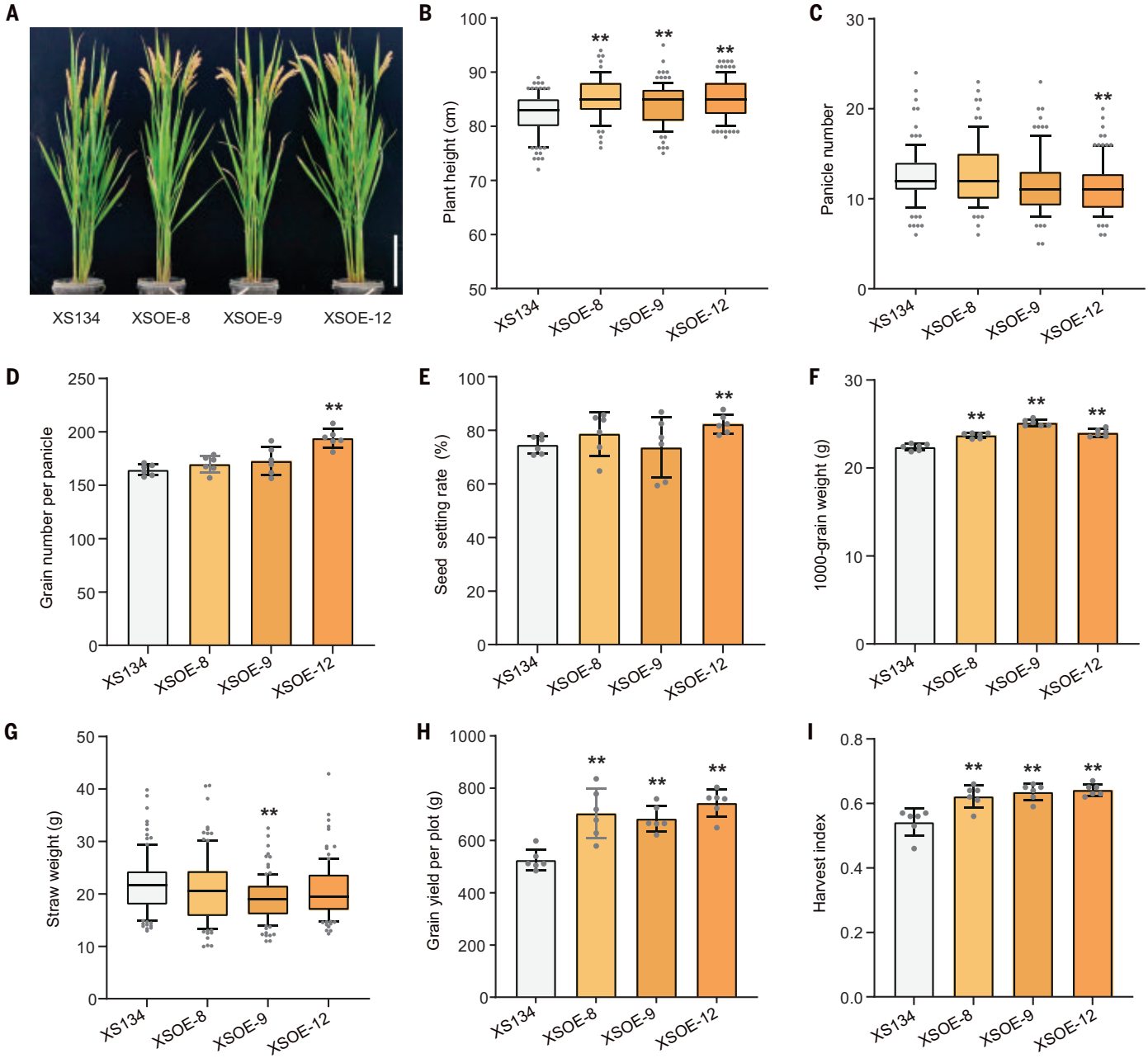


Fig. 5. OsDREB1C confers yield gains in an elite rice germplasm. (**A**) Whole-plant phenotypes of mature Xiushui134 (XS134) and XS134-*OsDREB1C*-OE plants (XSOE-8/9/12) grown in Hangzhou in 2020. Scale bar, 20 cm. (**B** to **I**) Yield parameters of XS134 and XS-OE plants grown in Hangzhou in 2021, including plant height (B), panicle number (C), grain number per panicle (D), seed setting rate (E), 1000-grain weight (F), straw weight (G), grain yield per plot (H), and harvest index (I). The box plots in panels (B), (C), and (G) show the median (horizontal line) and individual values (black dots) (n = 100 biological replicates). Data in F(D) to (I) except (G) are presented as means ± SD (n = 6 plots). *P < 0.05, **P < 0.01 compared with XS134, Student's t test.

assay (EMSA) confirmed that the DRE/CRT elements are necessary for OsDREB1C binding (Fig. 6F and fig. S22). Moreover, luciferase-based transient transactivation assays verified that OsDREB1C activates the expression of *OsRBCS3*, *OsNR2*, *OsNRT2.4*, *OsNRT1.1B*, and *OsFTL1* (Fig. 6G). Gene expression analyses revealed that the mRNA levels of *OsRBCS3*, *OsNR2*, *OsNRT2.4*, *OsNRT1.1B*, and *OsFTL1* correlated with *OsDREB1C* levels, in that the expression levels were increased in *OsDREB1C*-OE plants and decreased in *OsDREB1C*-KO plants (fig. S21B). Taken together, these results suggest that OsDREB1C can activate gene expression directly by binding to the promoter of *OsRBCS3* and to the exons of *OsNR2*, *OsNRT2.4*, *OsNRT1.1B*, and *OsFTL1*.

To clarify whether *OsFTL1* induction is responsible for the early-flowering phenotype of *OsDREB1C*-OE plants, we generated *OsFTL1* overexpression lines in the rice cultivar Nipponbare (fig. S23, A and E). The heading time of *OsFTL1*-OE plants was drastically shortened, ranging from 45 to 47 days, whereas that of WT plants was 116 to 118 days under the long-day photoperiod in Beijing (fig. S23, B and D). Under the short-day photoperiod in Hainan, *OsFTL1*-OE plants flowered 10 to 13 days earlier (fig. S23C). These results are consistent with a previous study on floral induction in transgenic plants grown in culture vessels (*26*). *OsFTL1*-OE plants were dwarfed and exhibited reduced grain yields (fig. S23F), probably because of their

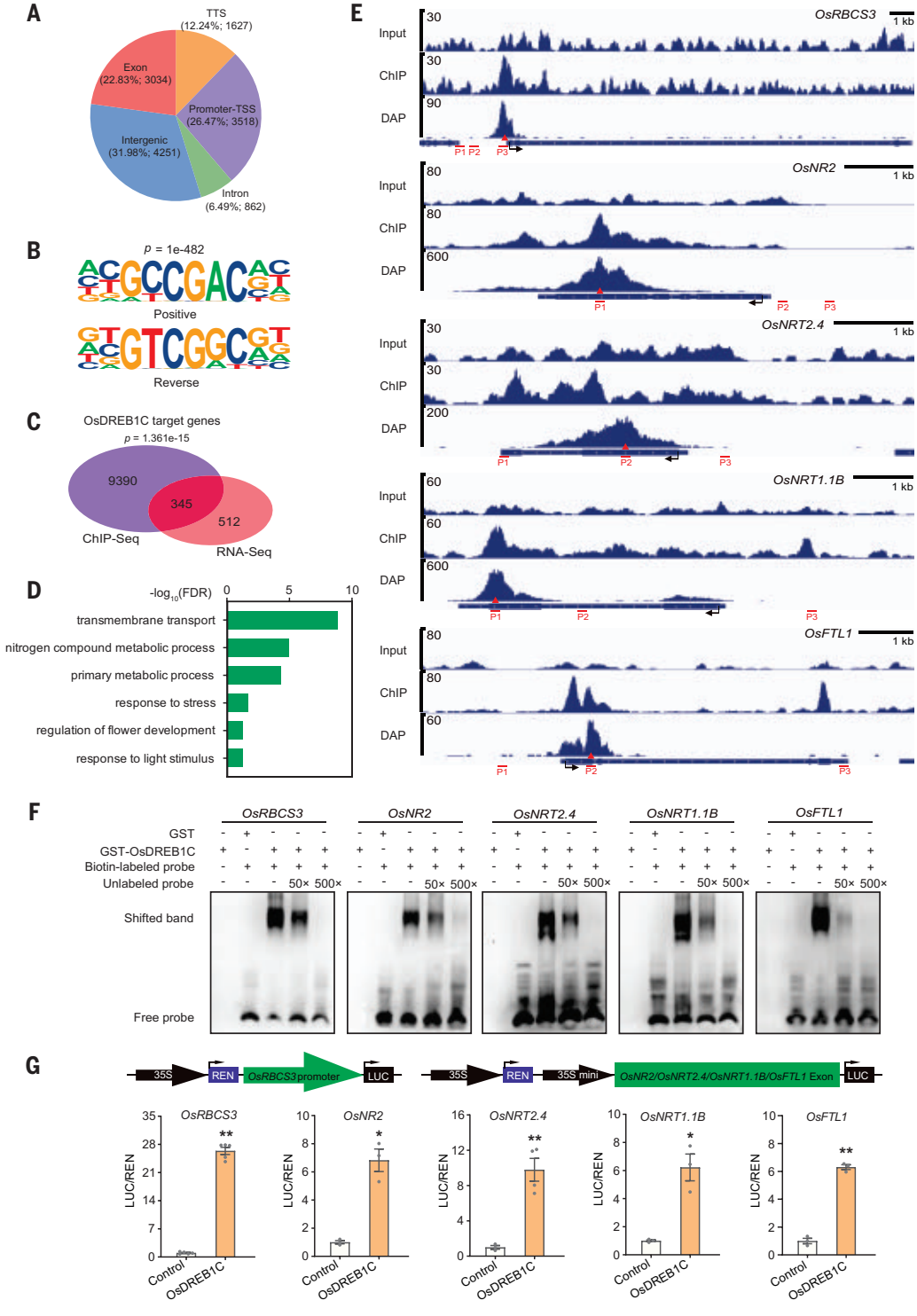


Fig. 6. OsDREB1C induces transcription of photosynthesis, nitrogen utilization, and flowering-related genes. (**A**) Distribution of candidate OsDREB1C-binding regions across the rice genome as determined by ChIP-seq. TSS, transcription start site; TTS, transcription termination site. (**B**) Motif analysis using HOMER to identify core motifs enriched within the experimentally determined (by ChIP-seq) OsDREB1C-binding regions. (**C**) Venn diagram showing the overlap between putative OsDREB1C target genes identified by ChIP-seq and differentially up-regulated genes in *OsDREB1C*-OE1 relative to the WT as identified by RNA-seq. (**D**) GO enrichment analysis of the overlapping gene set in (B). (**E**) OsDREB1C preferentially binds to the *OsRBCS3* promoter and to exons of *OsNR2*, *OsNRT2.4*, *OsNRT1.1B*, and *OsFTL1*, as validated by both ChIP-qPCR and DAP-seq. The diagrams in (E) depict the putative promoter region of *OsRBCS3* and exons of the *OsNR2*, *OsNRT2.4*, *OsNRT1.1B*, and *OsFTL1* genes. P1, P2, and P3 indicate primers used in the ChIP-qPCR experiments for the five examined loci shown in fig. S21. There are DRE/CRT elements within P3 of *OsRBCS3*, P1 of *OsNR2* and *OsNRT1.1B*, and P2 of *OsNRT2.4* and *OsFTL1*. (**F**) EMSA data confirming that the GST-OsDREB1C protein binds to promoters and exons containing DRE/CRT elements. (**G**) OsDREB1C activates transcription from the *OsRBCS3* promoter and from *OsNR2*, *OsNRT1.1B*, *OsNRT2.4*, and *OsFTL1* exon-luciferase fusion constructs in transient transactivation assays. Shown are relative ratios of the transcriptional activities conferred by OsDREB1C expression to the empty vector control. Asterisks indicate significant differences between the control and OsDREB1C. LUC/REN, ratio of firefly luciferase to *Renilla* luciferase activity. Data are presented as means ± SD ($n \geq 3$ biological replicates). *$P < 0.05$, **$P < 0.01$ compared with WT, Student's t test.

shortened vegetative phase (and the insufficient buildup of resources for allocation to seeds).

OsDREB1C effects in wheat and Arabidopsis

To determine whether the function of DREB1C is conserved in other plant species, we generated transgenic wheat and *Arabidopsis* plants overexpressing *OsDREB1C*. Analysis of the transgenic lines showed that *OsDREB1C* overexpression in the wheat cultivar Fielder also improved photosynthetic capacity, reduced the time to flowering by 3 to 6 days, and conferred increased grain yield per plant by 17.2 to 22.6% in the field and by 18.6 to 23.5% in the greenhouse (Fig. 7, A to F, and fig. S24). Likewise, transgenic *Arabidopsis* lines overexpressing *OsDREB1C* had more and bigger leaves, and flowered up to 4 days earlier than the WT (Fig. 7, G to J). The biomass yield was increased by 14.2 to 35.8% in the *OsDREB1C*-OE *Arabidopsis* plants (Fig. 7, K and L).

DISCUSSION

Transcription factors of the DREB subfamily belong to the AP2/ERF family and have been

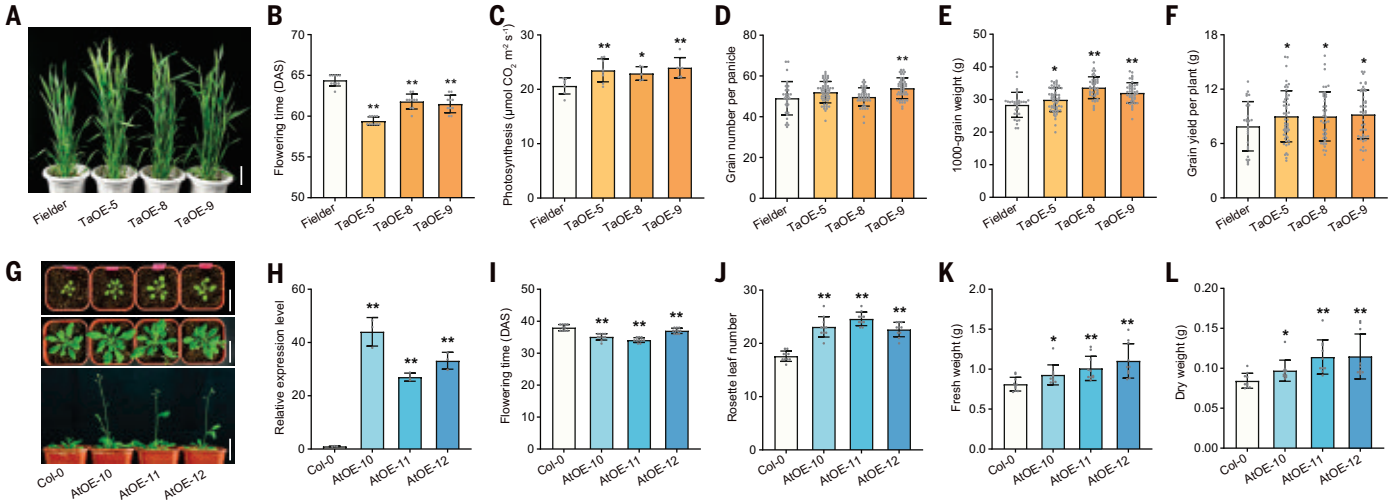


Fig. 7. OsDREB1C increases flowering, photosynthetic capacity, and yields in wheat and *Arabidopsis*. (**A**) Early flowering phenotype of *OsDREB1C*-OE field-grown wheat plants (*pUBI*::*OsDREB1C*, TaOE-5/8/9) compared with WT plants (cv. Fielder) at the booting stages. Scale bar, 10 cm. (**B** and **C**) Flowering time (B) and photosynthesis rate (C) of Fielder and *OsDREB1C*-OE wheat plants at the heading stage grown in the field in Beijing in 2021. Data are presented as means ± SD (n > 5 biological replicates). (**D** to **F**) Grain number per panicle (D), 1000-seed weight (E), and grain yield per plant (F) of Fielder and *OsDREB1C*-OE wheat plants grown in the field in Beijing in 2021. Data are presented as means ± SD (n > 30 biological replicates). *P < 0.05, **P < 0.01 compared with Fielder, Student's t test. (**G**) Phenotype of WT (Col-0) and *OsDREB1C*-OE *Arabidopsis* (*p35S*::*OsDREB1C-GFP*, AtOE-10/11/12) plants at the seedling and flowering stages. Note the early flowering of all three overexpression lines. Plants were grown in short-day conditions (8 hours light/16 hours dark) in a growth chamber for 2 weeks and then transferred to long-day conditions (16 hours light/8 hours dark). Scale bar, 4 cm. (**H**) *OsDREB1C* expression levels in *OsDREB1C*-OE *Arabidopsis* plants. Data are presented as means ± SD (n = 3 biological replicates). (**I** to **L**) Flowering time (I), rosette leaf number (J), fresh weight (K), and dry weight (L) of WT (Col-0) and *OsDREB1C*-OE *Arabidopsis* plants. Data are presented as means ± SD (n = 10 biological replicates). *P < 0.05, **P < 0.01 compared with Col-0, Student's t test.

demonstrated to activate multiple downstream genes in response to abiotic stresses such as drought, salinity, and freezing in various seed plants (*27*, *28*). The DREB-binding DRE/CRT *cis*-element (GCCGAC motif) is present in the promoter of many stress-inducible genes (*29*). However, the previous work has also revealed a trade-off between growth and stress tolerance, in that constitutive overexpression of *DREB1* genes in *Arabidopsis* and rice, although conferring improved stress tolerance, often leads to growth retardation and yield penalties (*27*, *29*–*31*). In the course of this work, we have shown that the inverse relationship between stress tolerance and yield can be uncoupled and even reversed. Overexpression of *OsDREB1C* in rice plants resulted in substantial yield increases of 41.3 to 68.3%, and these yield gains were accompanied by a shortened vegetative phase, in that the overexpression plants flowered much earlier than the WT. Although "high-yielding" and "early-maturing" have long been seen as conflicting traits in crop breeding, a report has shown that the long noncoding RNA *Ef-cd* shortens maturity duration without incurring a yield penalty (*32*). Similarly, overexpression of the nitrate transporter gene *OsNRT1.1A* confers both high yield and early maturation (*33*).

An unsolved problem pertinent to both agricultural productivity and the environmental footprint of current agricultural practices is the requirement for a high level of nitrogen fertilizer to attain high yields. This is mainly because of the low NUE of most major crop varieties (*34*). In addition to its negative environmental impact, excessive application of nitrogen fertilizer has other undesired effects, including delayed flowering, extended growth duration, and reduced yield potential (*35*). In this study, we have shown that by engineering the expression of a transcription factor that controls and coordinates photosynthesis, nitrogen utilization, and flowering time without affecting known genes involved in high yield and early flowering (fig. S25), it is possible to achieve enhanced growth and increased yields while at the same time improving the efficiency of nitrogen utilization. Thus, our work demonstrates that three key agricultural traits can be improved simultaneously by *OsDREB1C* overexpression: yield, NUE, and flowering time. We provide several lines of physiological and molecular evidence in support of the assumption that OsDREB1C confers accelerated vegetative growth and biomass accumulation before heading by (i) enhancing photosynthetic capacity through *OsRBCS3*; (ii) enhancing nitrogen uptake and transport through expression of *OsNRT1.1B*, *OsNRT2.4*, and *OsNR2*; and (iii) promoting early flowering through *OsFTL1*. We propose that efficient subsequent allocation of assimilated carbohydrates and nitrogen from leaves to the panicle further contributes to the elevated yield, presumably through coordinated regulation of several target genes of OsDREB1C, including amino acid and ammonium transporters (figs. S26 and S27). OsDREB1C binds to the exon regions rather than the promoters of four of the five verified target genes. However, the preferential binding of transcription factors to intragenic regions (exons and/or introns) of target genes has been demonstrated in a number of previous studies (*10*, *36*–*38*).

Currently, the relative rates of yield increase achieved by plant breeding are declining and have fallen below 1% per year for most cereal crops (*39*). In view of this trend and the need to double the world's food production by 2050 despite reduced availability of arable land and the challenges of climate change, the very large yield increases achieved in the field by engineering the expression of a single transcriptional regulator gene are unprecedented. Our findings suggest that after centuries of breeding for yield, there is still potential for substantial leaps in the yields of the world's main staple crops.

In the present study, overexpression of *OsDREB1C* was achieved using transgenic technologies. Alternatively, genome-editing technologies could be used to achieve *OsDREB1C* overexpression, for example, by using base editors to introduce expression-enhancing point mutations into the promoter region of the *OsDREB1C* gene (*40*, *41*), thus creating transgene-free, high-yielding varieties. Also, the existing natural

variation of OsDREB1C in rice provides a genetic resource that can be readily tapped. Finally, the conserved function of OsDREB1C in seed plants offers the potential to substantially improve biomass and yield in other crops. In summary, our data suggest OsDREB1C engineering as a generally applicable strategy to increase crop yields, with the added benefits of shortening the growth duration and lowering the environmental footprint of agriculture.

Methods summary

The candidate gene *OsDREB1C* was identified by a screen of 118 transcription factors that are related to C_4 photosynthesis, induced by light, and highly expressed under low-nitrogen conditions. The function of *OsDREB1C* was investigated by transgenic overexpression and CRISPR-Cas9–based gene-editing approaches. *Agrobacterium*-mediated transformation was used to introduce the constructed vectors into rice (Nipponbare and Xiushui134) and wheat (Fielder), and the floral dip method was used to transform *Arabidopsis* (Col-0). Gene expression levels in the overexpression and KO lines were evaluated by quantitative reverse transcription PCR (qRT-PCR).

The growth rates of hydroponically grown rice seedlings were monitored under standard growth conditions. Etiolated growth was assessed in constant darkness. Photosynthesis rates were measured with a LICOR-6400XT gas exchange system in field-grown plants. Accumulation levels of photosynthesis-related proteins were assessed by immunoblotting. RuBisCO content was quantified by immunoblot assays, and RuBisCO activity was determined as the rate of CO_2 fixation on RuBP using spectrophotometry. Sugar and starch contents were measured using enzymatic digestion methods. Nitrogen uptake and transport activity were evaluated with the ^{15}N-labeling assay. Carbon and nitrogen contents were assessed with an IsoPrime 100 analyzer. NUE was calculated as the ratio of grain yield to applied nitrogen fertilizer. Phytohormone contents were quantified using ultraperformance liquid chromatography quadropule ion trap mass spectrometry. Agronomic and yield traits were assessed in field experiments with randomized block design and performing three replicates for each experiment in Beijing, Hangzhou, and Hainan in four successive years (May 2018 to May 2022).

A sequence alignment of OsDREB1C orthologs was constructed and phylogenetic analysis was performed using ESPript 3.0 and MEGA 7, respectively. Genetic variation and haplotype association analyses were performed by variant cell format tools using a general linear model implemented in TASSEL 5.0. Subcellular localization of OsDREB1C was assessed by transient expression assays in rice protoplasts and tobacco leaves and in stable transgenic *Arabidopsis* plants expressing OsDREB1C-GFP. The binding ability of OsDREB1C to the predicted core promoter motifs was tested with yeast one-hybrid assays. The transactivation capacity of OsDREB1C was evaluated by transient transactivation assays.

To identify the genome-wide binding sites of OsDREB1C, ChIP-seq experiments were performed with rice protoplasts transiently expressing the OsDREB1C-GFP fusion protein, and RNA-seq analysis with WT and *OsDREB1C*-OE plants grown in the field. Up-regulated genes in *OsDREB1C*-OE plants associated with OsDREB1C-binding peaks were extracted as putative OsDREB1C targets, and GO enrichment analysis was conducted to identify the associated biological processes. ChIP-qPCR using transgenic plants expressing an OsDREB1C-GFP fusion protein and DAP-seq using expressed HALOTag-OsDREB1C fusion proteins in vitro were performed to verify the binding of OsDREB1C to target genes and test for target gene activation. Direct binding of OsDREB1C to target sequences was examined by EMSA. The transcriptional activation of targets genes by OsDREB1C was tested by dual luciferase reporter assays and qRT-PCR.

Details for experimental procedures are provided in the supplementary materials.

REFERENCES AND NOTES

1. Food and Agriculture Organization of the United Nations, "Hunger and food insecurity" (FAO, 2020); https://www.fao.org/hunger/en/.
2. D. Tilman, C. Balzer, J. Hill, B. L. Befort, Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20260–20264 (2011). doi: 10.1073/pnas.1116437108; pmid: 22106295
3. Food and Agriculture Organization of the United Nations, "The future of food and agriculture: Trends and challenges" (FAO, 2017); https://www.fao.org/3/i6583e/i6583e.pdf.
4. J. Bailey-Serres, J. E. Parker, E. A. Ainsworth, G. E. D. Oldroyd, J. I. Schroeder, Genetic strategies for improving crop yields. *Nature* **575**, 109–118 (2019). doi: 10.1038/s41586-019-1679-0; pmid: 31695205
5. G. M. Coruzzi, L. Zhou, Carbon and nitrogen sensing and signaling in plants: Emerging 'matrix effects'. *Curr. Opin. Plant Biol.* **4**, 247–253 (2001). doi: 10.1016/S1369-5266(00)00168-0; pmid: 11312136
6. A. Nunes-Nesi, A. R. Fernie, M. Stitt, Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions. *Mol. Plant* **3**, 973–996 (2010). doi: 10.1093/mp/ssq049; pmid: 20926550
7. S. Li *et al*., Modulating plant growth-metabolism coordination for sustainable agriculture. *Nature* **560**, 595–600 (2018). doi: 10.1038/s41586-018-0415-5; pmid: 30111841
8. K. Wu *et al*., Enhanced sustainable green revolution yield via nitrogen-responsive chromatin modulation in rice. *Science* **367**, eaaz2046 (2020). doi: 10.1126/science.aaz2046; pmid: 32029600
9. K. Kaufmann, C. A. Airoldi, Master regulatory transcription factors in plant development: A blooming perspective. *Methods Mol. Biol.* **1830**, 3–22 (2018). doi: 10.1007/978-1-4939-8657-6_1; pmid: 30043361
10. S. J. Burgess *et al*., Genome-wide transcription factor binding in leaves from C_3 and C_4 grasses. *Plant Cell* **31**, 2297–2314 (2019). doi: 10.1105/tpc.19.00078; pmid: 31427470
11. P. Cubas, N. Lauter, J. Doebley, E. Coen, The TCP domain: A motif found in proteins regulating plant growth and development. *Plant J.* **18**, 215–222 (1999). doi: 10.1046/j.1365-313X.1999.00444.x; pmid: 10363373
12. J. Doebley, A. Stec, L. Hubbard, The evolution of apical dominance in maize. *Nature* **386**, 485–488 (1997). doi: 10.1038/386485a0; pmid: 9087405
13. Y. Jiao *et al*., Regulation of *OsSPL14* by OsmiR156 defines ideal plant architecture in rice. *Nat. Genet.* **42**, 541–544 (2010). doi: 10.1038/ng.591; pmid: 20495565
14. J. Wang *et al*., A single transcription factor promotes both yield and immunity in rice. *Science* **361**, 1026–1028 (2018). doi: 10.1126/science.aat7675; pmid: 30190406
15. M. M. Ambavaram *et al*., Coordinated regulation of photosynthesis in rice increases yield and tolerance to environmental stress. *Nat. Commun.* **5**, 5302 (2014). doi: 10.1038/ncomms6302; pmid: 25358745
16. S. J. H. Kuijt *et al*., Interaction between the *GROWTH-REGULATING FACTOR* and *KNOTTED1-LIKE HOMEOBOX* families of transcription factors. *Plant Physiol.* **164**, 1952–1966 (2014). doi: 10.1104/pp.113.222836; pmid: 24532604
17. L. Wang *et al*., Comparative analyses of C_4 and C_3 photosynthesis in developing leaves of maize and rice. *Nat. Biotechnol.* **32**, 1158–1165 (2014). doi: 10.1038/nbt.3019; pmid: 25306245
18. W. Xin *et al*., An integrated analysis of the rice transcriptome and metabolome reveals root growth regulation mechanisms in response to nitrogen availability. *Int. J. Mol. Sci.* **20**, 5893 (2019). doi: 10.3390/ijms20235893; pmid: 31771277
19. S. Kojima *et al*., *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. *Plant Cell Physiol.* **43**, 1096–1105 (2002). doi: 10.1093/pcp/pcf156; pmid: 12407188
20. R. Komiya, A. Ikegami, S. Tamaki, S. Yokoi, K. Shimamoto, *Hd3a* and *RFT1* are essential for flowering in rice. *Development* **135**, 767–774 (2008). doi: 10.1242/dev.008631; pmid: 18223202
21. X. Li *et al*., Analysis of genetic architecture and favorable allele usage of agronomic traits in a large collection of Chinese rice accessions. *Sci. China Life Sci.* **63**, 1688–1702 (2020). doi: 10.1007/s11427-019-1682-6; pmid: 32303966
22. Y. Sakuma *et al*., Functional analysis of an *Arabidopsis* transcription factor, DREB2A, involved in drought-responsive gene expression. *Plant Cell* **18**, 1292–1309 (2006). doi: 10.1105/tpc.105.035881; pmid: 16617101
23. I. Kardailsky *et al*., Activation tagging of the floral inducer FT. *Science* **286**, 1962–1965 (1999). doi: 10.1126/science.286.5446.1962; pmid: 10583961
24. M. J. Yanovsky, S. A. Kay, Molecular basis of seasonal time measurement in *Arabidopsis*. *Nature* **419**, 308–312 (2002). doi: 10.1038/nature00996; pmid: 12239570
25. M. A. Parry, A. J. Keys, P. J. Madgwick, A. E. Carmo-Silva, P. J. Andralojc, Rubisco regulation: A role for inhibitors. *J. Exp. Bot.* **59**, 1569–1580 (2008). doi: 10.1093/jxb/ern084; pmid: 18436543
26. T. Izawa *et al*., Phytochrome mediates the external light signal to repress FT orthologs in photoperiodic flowering of rice. *Genes Dev.* **16**, 2006–2020 (2002). doi: 10.1101/gad.999202; pmid: 12154129
27. Y. Ito *et al*., Functional analysis of rice DREB1/CBF-type transcription factors involved in cold-responsive gene expression in transgenic rice. *Plant Cell Physiol.* **47**, 141–153 (2006). doi: 10.1093/pcp/pci230; pmid: 16284406
28. Q. Liu *et al*., Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. *Plant Cell* **10**, 1391–1406 (1998). doi: 10.1105/tpc.10.8.1391; pmid: 9707537
29. J. G. Dubouzet *et al*., *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *Plant J.* **33**, 751–763 (2003). doi: 10.1046/j.1365-313X.2003.01661.x; pmid: 12609047
30. M. Kudo *et al*., A gene-stacking approach to overcome the trade-off between drought stress tolerance and growth in Arabidopsis. *Plant J.* **97**, 240–256 (2019). doi: 10.1111/tpj.14110; pmid: 30285298
31. Y. Shi, Y. Ding, S. Yang, Molecular regulation of CBF signaling in cold acclimation. *Trends Plant Sci.* **23**, 623–637 (2018). doi: 10.1016/j.tplants.2018.04.002; pmid: 29735429
32. J. Fang *et al*., *Ef-cd* locus shortens rice maturity duration without yield penalty. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 18717–18722 (2019). doi: 10.1073/pnas.1815030116; pmid: 31451662
33. W. Wang *et al*., Expression of the nitrate transporter gene *OsNRT1.1A/OsNPF6.3* confers high yield and early maturation in rice. *Plant Cell* **30**, 638–651 (2018). doi: 10.1105/tpc.17.00809; pmid: 29475937
34. S. M. Swarbreck *et al*., A roadmap for lowering crop nitrogen requirement. *Trends Plant Sci.* **24**, 892–904 (2019). doi: 10.1016/j.tplants.2019.06.006; pmid: 31285127

35. H. Li, B. Hu, C. Chu, Nitrogen use efficiency in crops: Lessons from *Arabidopsis* and rice. *J. Exp. Bot.* **68**, 2477–2488 (2017). doi: 10.1093/jxb/erx101; pmid: 28419301
36. Z. Dong *et al*., Ideal crop plant architecture is mediated by *tassels replace upper ears1*, a BTB/POZ ankyrin repeat gene directly targeted by TEOSINTE BRANCHED1. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E8656–E8664 (2017). doi: 10.1073/pnas.1714960114; pmid: 28973898
37. E. González-Grandío *et al*., Abscisic acid signaling is controlled by a BRANCHED1/HD-ZIP I cascade in *Arabidopsis* axillary buds. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E245–E254 (2017). doi: 10.1073/pnas.1613199114; pmid: 28028241
38. X. Yang *et al*., Regulation of plant architecture by a new histone acetyltransferase targeting gene bodies. *Nat. Plants* **6**, 809–822 (2020). doi: 10.1038/s41477-020-0715-2; pmid: 32665652
39. R. A. T. Fischer, G. O. Edmeades, Breeding and cereal yield progress. *Crop Sci.* **50**, S-85–S-98 (2010). doi: 10.2135/cropsci2009.10.0564
40. A. V. Anzalone *et al*., Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019). doi: 10.1038/s41586-019-1711-4; pmid: 31634902
41. J. Tan, F. Zhang, D. Karcher, R. Bock, Expanding the genome-targeting scope and the site selectivity of high-precision base editors. *Nat. Commun.* **11**, 629 (2020). doi: 10.1038/s41467-020-14465-z; pmid: 32005820

ACKNOWLEDGMENTS

We thank X. Li (Institute of Crop Sciences, CAAS) for help with drawing the working model and X. Ye and K. Wang (Institute of Crop Sciences, CAAS) for help with wheat transformation. **Funding:** This research was supported by the National Key Research and Development Program of China (grants 2016YFD0300100 and 2016YFD0300102). W.Z. was supported by the Innovation Program of the Chinese Academy of Agricultural Sciences and the Elite Youth Program of the Chinese Academy of Agricultural Science. **Author contributions:** W.Z., S.W., and X.L. conceived and designed the experiments. S.W. and X.L. performed most of the experiments. H.Z., X.L., and S.W. produced the transgenic plants. X.Y., Y.Z., J.L., Y.Y., and F.D. characterized the phenotypes of transgenic plants. D.W. and S.C. performed field experiments with Xiushui134 rice transgenic plants. Z.L., H.P., and S.W. analyzed DAP-seq and RNA-seq results. Y.Z. and P.Y. performed the ChIP-qPCR experiment. S.W. and Z.L. analyzed the ChIP-seq results. L.S. and C.Z. performed haplotype analysis. S.W., X.L., and W.Z. wrote the manuscript. R.B., J.H., W.P., Q.Q., and M.Z. critically commented on and edited the manuscript. All authors discussed and commented on the manuscript. **Competing interests:** The authors declare no competing interests. Patent applications related to this work have been submitted by W.Z., S.W., and X.L. **Data and materials availability:** Raw sequence data generated during this study have been deposited in the National Center for Biotechnology Information (NCBI) BioProject database under accession number PRJNA724935 for RNA-seq, PRJNA841272 for ChIP-seq, and PRJNA841281 for DAP-seq. All other data are available in the main text or the supplementary materials. Requests for materials should be addressed to W.Z.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi8455
Materials and Methods
Figs. S1 to S27
Tables S1 to S10
References (*42*–*70*)
MDAR Reproducibility Checklist

Submitted 10 May 2021; resubmitted 29 March 2022
Accepted 26 May 2022
10.1126/science.abi8455

RESEARCH ARTICLES

PROTEIN DESIGN

Scaffolding protein functional sites using deep learning

Jue Wang[1,2]†, Sidney Lisanza[1,2,3]†, David Juergens[1,2,4]†, Doug Tischer[1,2]†, Joseph L. Watson[1,2]†, Karla M. Castro[5], Robert Ragotte[1,2], Amijai Saragovi[1,2], Lukas F. Milles[1,2], Minkyung Baek[1,2], Ivan Anishchenko[1,2], Wei Yang[1,2], Derrick R. Hicks[1,2], Marc Expòsit[1,2,4], Thomas Schlichthaerle[1,2], Jung-Ho Chun[1,2,3], Justas Dauparas[1,2], Nathaniel Bennett[1,2,4], Basile I. M. Wicky[1,2], Andrew Muenks[1,2], Frank DiMaio[1,2], Bruno Correia[5], Sergey Ovchinnikov[6,7]*, David Baker[1,2,8]*

The binding and catalytic functions of proteins are generally mediated by a small number of functional residues held in place by the overall protein structure. Here, we describe deep learning approaches for scaffolding such functional sites without needing to prespecify the fold or secondary structure of the scaffold. The first approach, "constrained hallucination," optimizes sequences such that their predicted structures contain the desired functional site. The second approach, "inpainting," starts from the functional site and fills in additional sequence and structure to create a viable protein scaffold in a single forward pass through a specifically trained RoseTTAFold network. We use these two methods to design candidate immunogens, receptor traps, metalloproteins, enzymes, and protein-binding proteins and validate the designs using a combination of in silico and experimental tests.

The biochemical functions of proteins are often carried out by a subset of residues that constitute a functional site—for example, an enzyme active site or a protein or small-molecule binding site—and hence the design of proteins with new functions can be divided into two steps. The first step is to identify functional site geometries and amino acid identities that produce the desired activity—for enzymes, this can be done using quantum chemistry calculations (*1*–*3*), and for protein binders, by fragment docking calculations (*4*, *5*). Alternatively, functional sites can be extracted from a native protein having the desired activity (*6*, *7*). Here, we focus on the second step: Given a functional site description from any source, design an amino acid sequence that folds up to a three-dimensional (3D) structure containing the site. Previous methods can scaffold functional sites made up of one or two contiguous chain segments (*6*–*10*), but, with the exception of helical bundles (*8*), these do not extend readily to more complex sites composed of three or more chain segments, and the generated backbones are not guaranteed to be designable (i.e., encodable by some amino acid sequence).

An ideal method for functional de novo protein design would (i) embed the functional site with minimal distortion in a designable scaffold protein; (ii) be applicable to arbitrary site geometries, searching over all possible scaffold topologies and secondary structure compositions for those optimal for harboring the specified site; and (iii) jointly generate backbone structure and amino acid sequence. We previously demonstrated that the trRosetta structure-prediction neural network (*11*) can be used to generate new proteins by maximizing the trRosetta output probability that a sequence folds to some (unspecified) 3D structure during Monte Carlo sampling in sequence space (*12*). We refer to this process as "hallucination," as it produces solutions that the network considers to be ideal proteins but that do not correspond to any known natural protein; crystal and nuclear magnetic resonance structures confirm that the hallucinated sequences fold to the hallucinated structures (*12*). trRosetta can also be used to design sequences that fold into a target backbone structure by carrying out sequence optimization using a structure recapitulation loss function that rewards similarity of the predicted structure to the target structure (*13*). Given this ability to design both sequence and structure, we reasoned that trRosetta could be adapted to tackle the functional site scaffolding problem.

[1]Department of Biochemistry, University of Washington, Seattle, WA 98105, USA. [2]Institute for Protein Design, University of Washington, Seattle, WA 98105, USA. [3]Graduate Program in Biological Physics, Structure and Design, University of Washington, Seattle, WA 98105, USA. [4]Molecular Engineering Graduate Program, University of Washington, Seattle, WA 98105, USA. [5]Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland. [6]FAS Division of Science, Harvard University, Cambridge, MA 02138, USA. [7]John Harvard Distinguished Science Fellowship Program, Harvard University, Cambridge, MA 02138, USA. [8]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98105, USA.
*Corresponding author. Email: dabaker@uw.edu (D.B.); so@fas.harvard.edu (S.O.)
†These authors contributed equally to this work.

Partially constrained hallucination using a multiobjective loss function

To extend existing trRosetta-based design methods to scaffold functional sites (Fig. 1A), we optimized amino acid sequences for folding to a structure containing the desired functional site using a composite loss function that combines the previously used hallucination loss with a motif reconstruction loss over the functional motif [rather than the entire structure, as in (*13*)] (Fig. 1B; see materials and methods in the supplementary materials). Although we succeeded in generating structures with segments closely recapitulating functional sites, Rosetta structure predictions suggested that the sequences poorly encoded the structures (fig. S1A), and hence we used Rosetta design calculations to generate more-optimal sequences (*14*). Several designs targeting programmed cell death ligand 1 (PD-L1) generated by constrained hallucination with binding motifs derived from programmed cell death protein 1 (PD-1) (table S1) (*15*), followed by Rosetta design, were found to have binding affinities in the mid-nanomolar range (fig. S1, B to E). Although this experimental validation is encouraging, the requirement for sequence design using Rosetta is inconsistent with the aim of jointly designing sequence and structure.

Following the development of RoseTTAFold (RF) (*16*), we found that it performed better than trRosetta in guiding protein design by functional site–constrained hallucination (fig. S1G), likely reflecting the better overall modeling of protein sequence-structure relationships (*16*). Constrained hallucination with RoseTTAFold has the further advantages that, because 3D coordinates are explicitly modeled (trRosetta only generates inter-residue distances and orientations), site recapitulation can be assessed at the coordinate level and additional problem-specific loss terms can be implemented in coordinate space that assess interactions with a target (fig. S2; materials and methods).

Generalized functional motif scaffolding by missing information recovery

While powerful and general, the constrained hallucination approach is compute-intensive, as a forward and backward pass through the network is required for each gradient descent step during sequence optimization. In the training of recent versions of RoseTTAFold, a subset of positions in the input multiple sequence alignment are masked, and the network is trained to recover this missing sequence information in addition to predicting structure. This ability to recover both sequence and structural information provides a second solution to the functional site scaffolding problem: Given a functional site description, a forward pass through the network can be used to complete, or "inpaint," both protein sequence and

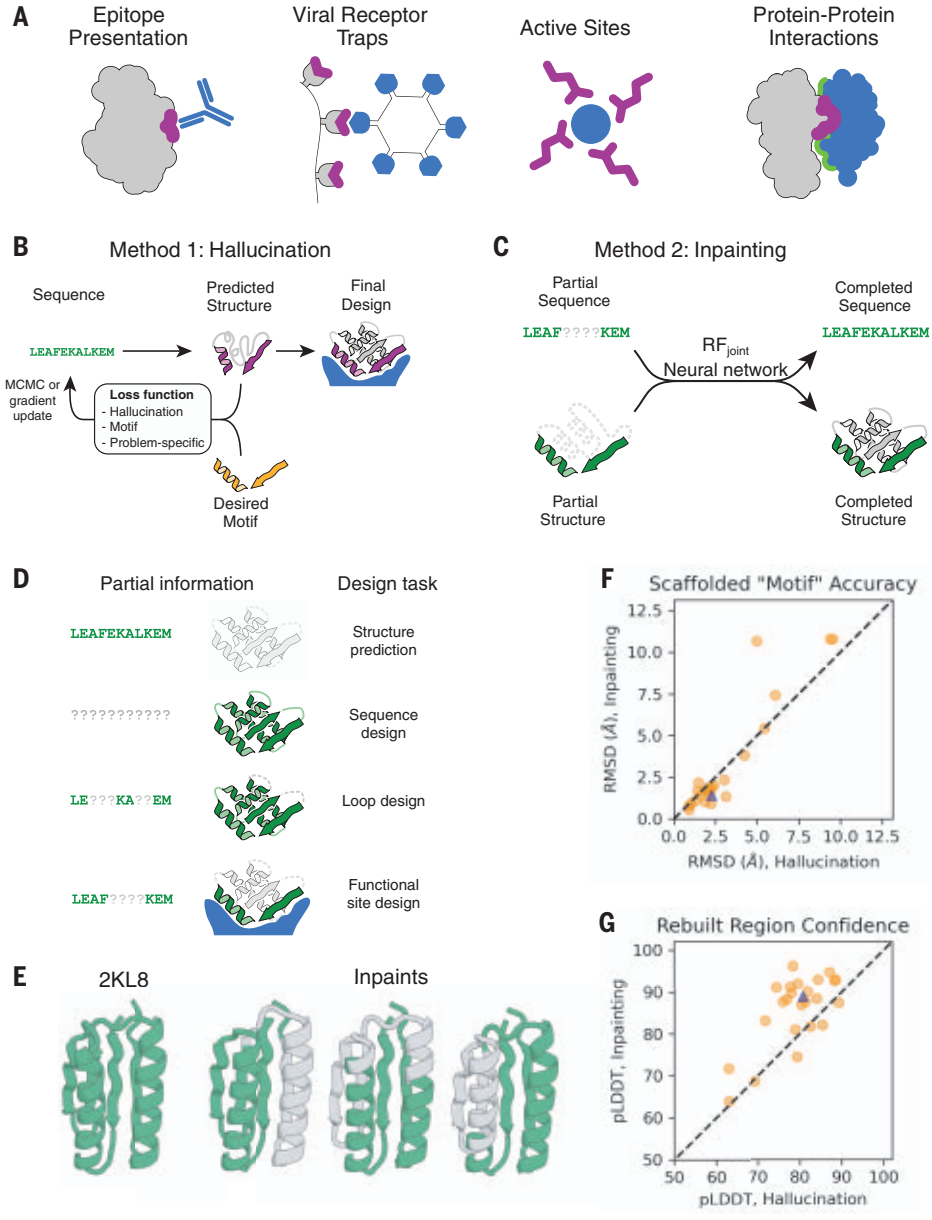


Fig. 1. Methods for protein function design. (**A**) Applications of functional-site scaffolding. (**B** and **C**) Design methods. (B) Constrained hallucination. At each iteration, a sequence is passed to the trRosetta or RoseTTAFold neural network, which predicts 3D coordinates and inter-residue distances and orientations (fig. S2). The predictions are scored by a loss function that rewards certainty of the predicted structure along with motif recapitulation and other task-specific functions. MCMC, Markov chain Monte Carlo. (C) Missing information recovery ("inpainting"). Partial sequence and/or structural information is input into a modified RoseTTAFold network (called RF_{joint}), and complete sequence and structure are output. (**D**) Protein design challenges formulated as missing information recovery problems. Question marks in column 1 indicate missing sequence information; gray cartoons in column 2, missing structural information. (**E**) RF_{joint} can simultaneously recover structure and sequence of a masked protein region. 2KL8 was fed into RF_{joint} with a continuous (length 30) window of sequence and structure masked out, with the network tasked with predicting the missing region of protein. Outputs (inpainted region in gray) closely resemble the original protein (2KL8, left) and are confidently predicted by AlphaFold (pLDDT/motif RMSD of models shown, from left to right: 91.6/0.91, 92.0/0.69, and 90.4/0.82). (**F** and **G**) Motif scaffolding benchmarking data comparing RF_{joint} with constrained hallucination. A set of 28 de novo designed proteins, published since RoseTTAFold was trained, were used. For each protein, 20 random masks of length 30 were generated, and RF_{joint} and hallucination were tasked with filling in the missing sequence and structure to "scaffold" the unmasked "motif." For this mask length, RF_{joint} typically modestly outperforms hallucination, both in terms of the RMSD of the unmasked protein (the "motif") to the original structure (F) and in AlphaFold confidence (pLDDT in the replaced region) (G). Circles represent average of 20 outputs for each of the benchmarking proteins. Triangle represents 2KL8. Colors in all panels: native functional motif, orange; hallucinated/inpainted scaffold, gray; constrained motif, purple; binding partner, blue; nonmasked region, green; and masked region, light-gray dotted lines.

structure in a masked region of protein (Fig. 1C; materials and methods). Here, the design challenge is formulated as an information recovery problem, analogous to the completion of a sentence given its first few words using language models (*17*) or the completion of corrupted images using inpainting (*18*). A wide variety of protein structure prediction and design challenges can be similarly formulated as missing information recovery problems (Fig. 1D). Although protein inpainting has been explored before (*19*, *20*), in this study we approach it using the power of a pretrained structure-prediction network.

We began from a RoseTTAFold (RF) model trained for structure prediction (*16*) and carried out further training on fixed-backbone sequence design in addition to the standard fixed-sequence structure prediction task to avoid model degradation (fig. S3; materials and methods). This model, denoted $RF_{implicit}$, was able to recover small, contiguous regions missing both sequence and structure (fig. S3). Encouraged by this result, we trained a model explicitly on inpainting segments with missing sequence and structure given the surrounding protein context, in addition to sequence design and structure prediction tasks (fig. S4A; materials and methods and algorithm S1). The resulting model was able to inpaint missing regions with high fidelity (Fig. 1E and fig. S4) and performed well at sequence design (32% native sequence recovery during training) and structure prediction (fig. S4C). We call this network RF_{joint} and use it to generate all inpainted designs below unless otherwise noted.

To evaluate in silico the quality of designs generated by our methods, we use the AlphaFold (AF) protein structure prediction network (*21*), which has high accuracy on de novo designed proteins (*22*) (fig. S7A). RF and AF have different architectures and were trained independently, and hence AF predictions can be regarded as a partially orthogonal in silico test of whether RF-designed sequences fold into the intended structures, analogous to traditional ab initio folding (*13*, *23*). We used AF to compare the ability of hallucination and

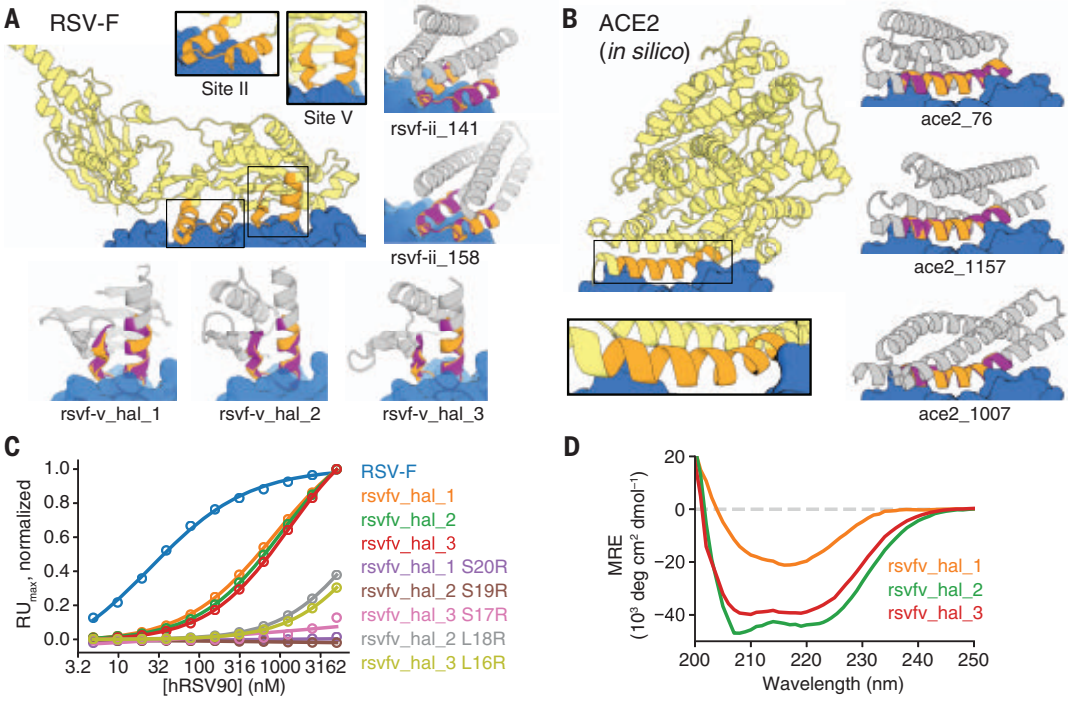


Fig. 2. Design of epitope scaffolds and receptor traps. (**A**) Design of proteins scaffolding immunogenic epitopes on RSV protein F (site II: PDB ID 3IXT chain P residues 254 to 277; site V: PDB ID 5TPN chain A residues 163 to 181). Comparisons of the RF hallucinated models to AF2 structure predictions from the design sequence are in fig. S9; here, because of space constraints, we show only the AF2 model (the two are very close in all cases). Here and in the following figures, we assess the extent of success in designing sequences that fold to structures harboring the desired motif through two metrics computed on the AF2 predictions: prediction confidence (AF pLDDT) and the accuracy of recapitulation of the original scaffolded motif (motif AF-RMSD). For RSV-F designs, these metrics are rsvf_ii_141 (85.0, 0.53 Å), rsvf_ii_158 (82.9, 0.51 Å), rsvf_ii_171 (88.4, 0.69 Å), rsvfv_hal_1 (82, 0.7 Å), rsvfv_hal_2 (88, 0.64 Å), and rsvfv_hal_3 (86, 0.65 Å). (**B**) Design of COVID-19 receptor trap based on ACE2 interface helix (PDB ID 6VW1 chain A residues 24 to 42). Design metrics: ace2_76 (89.1, 0.55 Å), ace2_1157 (80.4, 0.47 Å), and ace2_1007 (83.3, 0.57 Å). Colors: native protein scaffold, light yellow; native functional motif, orange; hallucinated scaffold, gray; hallucinated motif, purple; and binding partner, blue. See table S2 for additional metrics on each design. (**C**) Normalized maximum surface plasmon resonance signal (response units) of purified RSV-F epitope scaffolds and point mutants at various concentrations of hRSV90 antibody, with sigmoid fits. RSV-F refers to purified trimeric native F protein. K_d values are as follows: RSV-F: 24 nM; rsvfv_hal_1: 0.9 μM; rsvfv_hal_2: 1.0 μM; rsvfv_hal_3: 1.3 μM. (**D**) Mean residue ellipticity (MRE) versus wavelength, from CD spectroscopy, for the three RSV-F site V hallucinations with binding activity.

inpainting to rebuild missing protein regions (Fig. 1, F and G, and fig. S5). Inpainting yielded solutions with more accurately predicted fixed regions ("AF-RMSD"; Fig. 1G and fig. S5B) and structures overall more confidently predicted from their amino acid sequences ("AF pLDDT"; Fig. 1F and fig. S5A) and required only 1 to 10 s per design on an NVIDIA RTX 2080 graphics processing unit (hallucination requires 5 to 20 min per design). However, hallucination gave better results when the missing region was large (fig. S5) and generated greater structural diversity (fig. S8; and see below).

In the following sections, we highlight the power of the constrained hallucination and inpainting methods by designing proteins containing a wide range of functional motifs (Figs. 2 to 5 and table S1). For almost all problems, we obtained designs that are closely recapitulated by AF with overall and motif (functional site) root mean square deviation (RMSD) of typically <2 and <1 Å, respectively, with high model confidence [predicted local distance difference test (pLDDT) > 80; table S2]; such recapitulation suggests that the designed sequences encode the designed structures [although it should be noted that AF has limited ability to predict protein stability (*24*) or mutational effects (*25*, *26*)]. More critically, we assessed the activities of the designs experimentally (with the exception of those labeled "in silico" in Figs. 2 to 5).

Designing immunogen candidates and receptor traps

The goal of immunogen design is to scaffold a native epitope recognized by a neutralizing antibody as accurately as possible in order to elicit antibodies binding the native protein upon immunization. Additional interactions with the antibody are undesirable because the aim is to elicit antibodies recognizing only the original antigen, and hence for hallucination, we add a repulsive loss term to penalize interactions with the antibody beyond those present in the scaffolded epitope (fig. S2; supplementary text). As a test case, we focused on respiratory syncytial virus F protein (RSV-F), which has several antigenic epitopes for which structures with neutralizing antibodies have been determined (*7*, *9*, *10*). We scaffolded RSV-F site II, a 24-residue helix-loop-helix motif that had previously been grafted successfully onto a three-helix bundle (*7*), as well as RSV-F site V, a 19-residue helix-loop-strand motif that has not yet been scaffolded successfully (*27*). We were able to hallucinate designs recapitulating both epitopes to sub-angstrom backbone RMSD in a variety of folds [Fig. 2A and fig. S9; structures and sequences for all designs below are given in data S1 and S2 and differ considerably from native proteins (table S2); RF hallucinated models and AF structure predictions are shown in figs. S9, S11, and S17; only the AF model is shown in the main figures]. Inpainting also generated scaffolds for RSV-F site V, with comparable quality but less diversity than the hallucinations (fig. S8).

We expressed 37 hallucinated RSV-F site V scaffolds with high AF pLDDT and low motif AF-RMSD in *Escherichia coli* and found that three bound the neutralizing antibody hRSV90 (*27*) with a dissociation constant (K_d) of 0.9 to 1.3 μM (Fig. 2C and fig. S11; materials and methods and supplementary text). The K_d for the RSVF trimer is lower (23 nM), but the interface is larger, encompassing both sites II and V (*27*). Mutation of either of two key epitope residues reduced or abolished binding of the designs, suggesting that they bind the target through the scaffolded motif (Fig. 2C and fig. S11A), and circular dichroism (CD) spectra were consistent with the designed scaffold structures for both the original hallucinations (Fig. 2D) and the epitope mutants (fig. S11C). Four of the inpainted designs bound hRSV90 by yeast display but were poorly expressed in *E. coli*

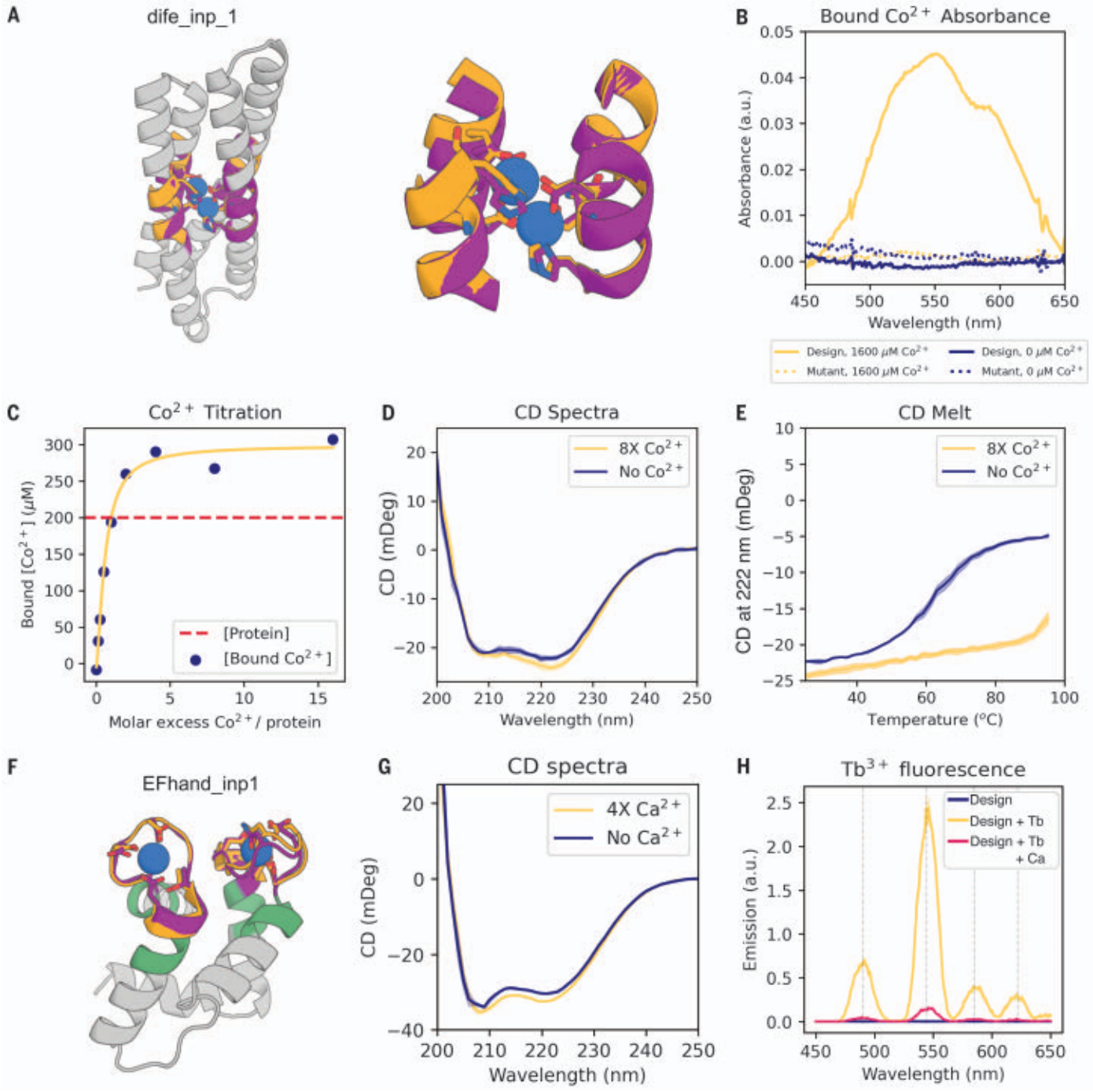


Fig. 3. Design of metal binding. (**A**) Scaffolding of di-iron binding site from *E. coli* cytochrome b1 (PDB ID 1BCF chain A residues 18 to 25, 27 to 54, 94 to 97, and 123 to 130) using inpainting. Colors: native protein scaffold, light yellow; native functional motif, orange; hallucinated scaffold, gray; hallucinated motif, purple; and bound metal, blue. (**B**) Absorbance spectra of difе_inp_1 (or mutant) in the presence (or absence) of an eight-fold molar excess of Co^{2+}. Peaks at 520, 555, and 600 nm, consistent with Co^{2+} binding to the scaffolded motif (*32*). In the mutant, the six coordinating residues [side chains shown in (A)] are mutated to alanine (E16A, E55A, H58A, E89A, H92A, E115A). Protein concentration: 200 μM. (**C**) difе_inp_1 Co^{2+} titration (protein concentration: 200 μM). Quantification of the absorbance at 550 nm, using a predicted extinction coefficient of 155 for Co^{2+} binding the motif (*32*), is consistent with both binding sites being recapitulated. (**D**) CD spectra of difе_inp_1 in the presence and absence of Co^{2+} are both consistent with the predicted helical structure. (**E**) Temperature dependence of difе_inp_1 CD signal in the presence and absence of Co^{2+}. Coordination of Co^{2+} in the core stabilizes the protein. Protein concentration: 6.7 μM; Co^{2+} concentration: 53.3 μM. (**F**) Inpainted design EFhand_inp_1 scaffolding the double EF-hand motif with input motif residues in purple, input nonmotif residues in green, and overlaid with the native motif from PDB ID 1PRW (orange). (**G**) CD spectra of EFhand_inp_1 incubated with and without $CaCl_2$ suggest stabilization of the protein upon binding calcium. (**H**) Tryptophan-enhanced terbium fluorescence spectra of EFhand_inp_1 suggests that the design binds terbium (*57*). Terbium binding signal is competed by 1 mM $CaCl_2$ (red). Design metrics (AF pLDDT, motif AF-RMSD): difе_inp_1 (92, 0.65 Å) and EFhand_inp1 (84, 0.7 Å).

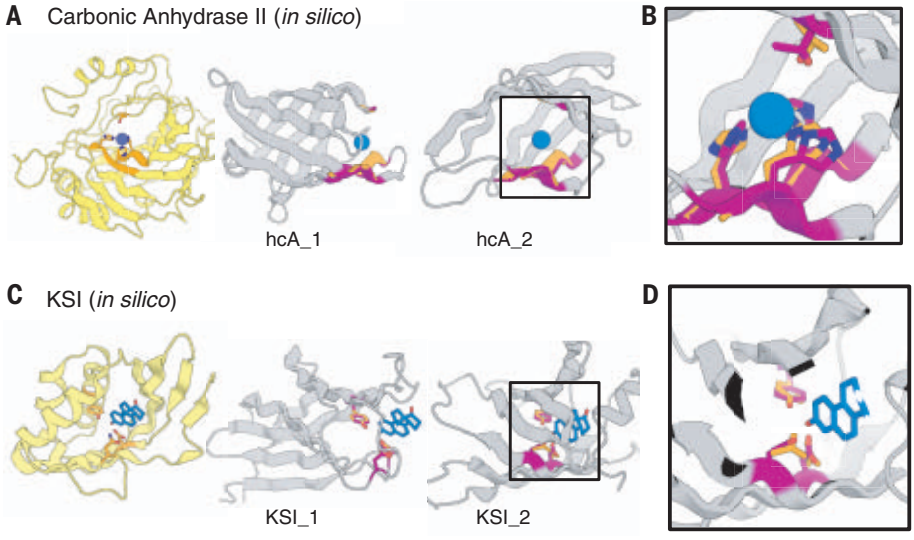


Fig. 4. In silico design of enzyme active sites. (**A** and **B**) Hallucinations using backbone description of site using RF. (**C** and **D**) Hallucination using side-chain description of site using AF2 augmented with trRosetta (materials and methods). (A) Carbonic anhydrase II active site (PDB ID 5YUI chain A residues 62 to 65, 93 to 97, and 118 to 120). (B) Δ^5-3-ketosteroid isomerase active site (PDB ID 1QJG chain A residues 14, 38, and 99). Colors: native protein scaffold, light yellow; native functional motif, orange; hallucinated scaffold, gray; hallucinated motif, purple; and bound metal, blue. [(B) and (D)] Zoomed-in view of designed active sites. Design metrics (AF pLDDT, motif AF-RMSD): hcA_1 (73, 1.04 Å), hcA_2 (71, 0.62 Å), KSI_1 (84, 0.30 Å Cβ), and KSI_2 (72, 0.53 Å Cβ).

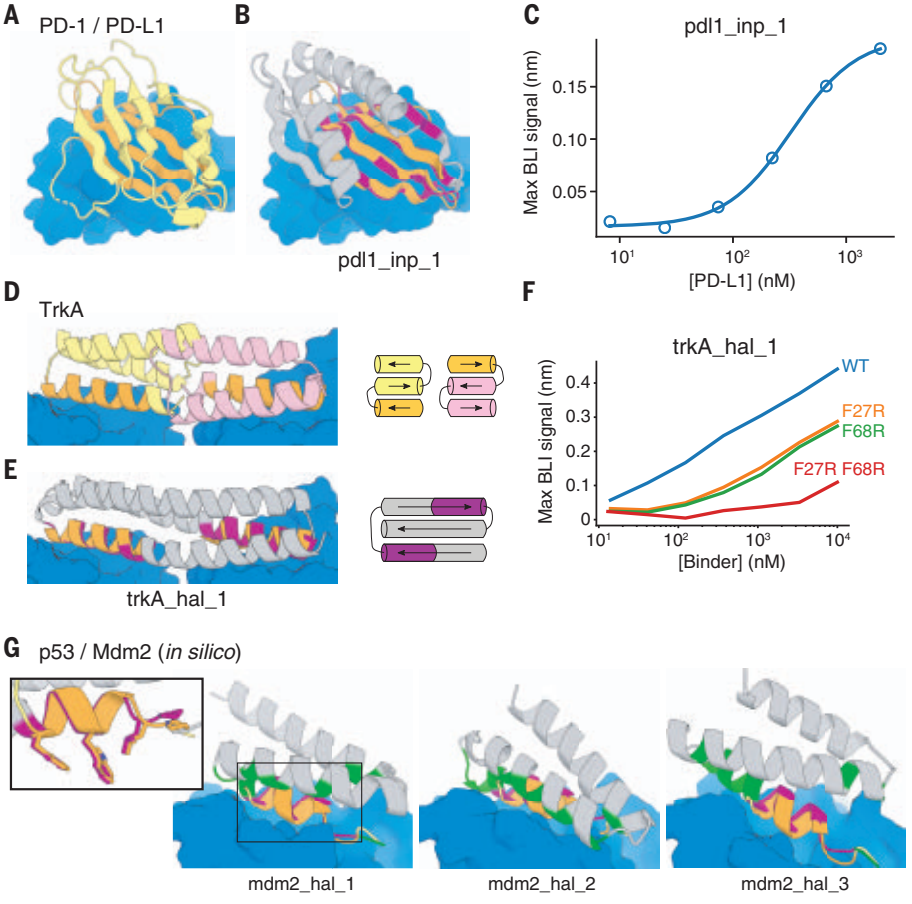


Fig. 5. Design of protein-binding proteins. Designs containing target-binding interfaces built around native-complex–derived binding motifs. Targets are in blue, native scaffolds in yellow or pink, native motifs in orange, designed scaffolds in gray, and designed motifs in purple. (**A**) Crystal structure of HAC PD-1 in complex with PD-L1. (**B**) Inpainted PD-L1 binder superimposed on PD-1 interface motif. (**C**) BLI binding signal versus PD-L1 concentration. K_d = 326 nM. (**D**) Crystal structure of previously designed TrkA minibinder in complex with TrkA, superimposed on TrkA receptor dimer. (**E**) Hallucinated bivalent TrkA binder. Protein topology diagrams are on the right. (**F**) BLI binding signal versus TrkA concentration; mutations at both scaffolded binding sites reduce TrkA binding. (**G**) Hallucinated Mdm2 binder designs superimposed on native p53 helix in complex with Mdm2 (see also fig. S17, D and E). New binding interactions (hallucinated residues within 5 Å of the target) are in green. (Inset) Overlay of mdm2_hal_1 and native p53 helix showing key side chains for binding.

(fig. S11, C to E). Overall, the designs provide a diverse set of promising starting points for further RSV-F epitope-based vaccine development.

We next applied hallucination to the in silico design of receptor traps that neutralize viruses by mimicking their natural binding targets and thus are inherently robust against mutational escape. We again augmented the loss function with a penalty on interactions beyond those in the native receptor to avoid opportunities for viral escape. As a test case, we scaffolded the helix of human angiotensin-converting enzyme 2 (hACE2) interacting with the receptor binding domain of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein (*28*). The hallucinated hACE2 mimetics have a diverse set of helical topologies, and AF structure predictions recapitulate the binding interface with sub-angstrom accuracy (Fig. 2B and fig. S9C).

Designing metal-coordinating proteins

Di-iron sites are important in biological systems for iron storage (*29*) and can mediate catalysis (*30*, *31*). We were able to recapitulate the di-iron site from *E. coli* bacterioferritin, composed of four parallel helical segments, to

sub-angstrom AF-RMSD using both inpainting (Fig. 3, A to E, and fig. S13) and hallucination (fig. S12; the hallucinations were not tested owing to buried polar residues; supplementary text). The designs had diverse helix connectivities and low structural similarity to the parent [figs. S13B and S12; template modeling (TM)–score 0.55 to 0.71 to PDB ID 1BCF_A]. We chose 96 inpainted designs to test experimentally and found that 76 had soluble expression, at least eight (see supplementary text) had a spectroscopic shift indicative of Co^{2+} binding (a proxy for iron binding) (*32*, *33*), and three (dife_inp_1, dife_inp_2, and dife_inp_3; Fig. 3B and fig. S13E) had CD spectra consistent with the designed fold (Fig. 3D and fig. S13F) and were stabilized by metal binding (Fig. 3E and fig. S13G). Mutation of the metal binding residues abolished binding (Fig. 3B and fig. S13E), and titration analysis of dife_inp_1 suggested that both metal binding sites were successfully scaffolded (Fig. 3C).

We next scaffolded the calcium-binding EF-hand motif (*34*), a 12-residue loop flanked by helices. Both constrained hallucination and inpainting readily generated scaffolds recapitulating either one or two EF-hand motifs to within 1.0 Å AF-RMSD of the native motif (Fig. 3F; fig. S14, A and B; and table S2). We chose 20 hallucinations and 55 inpaints to display on yeast and screen for calcium binding using tryptophan-enhanced terbium fluorescence (*35*). Six hallucinations and four inpaintings had fluorescence consistent with ion binding [fig. S14A; materials and methods; one of these proteins (EFhand_inp_2) was designed using $RF_{implicit}$ (supplementary text)]. The top hit from yeast, the inpainted EFhand_inp_1, purified from *E. coli* as a monomer (fig. S14C), had the expected CD spectrum (Fig. 3G) and a clear terbium binding signal (Fig. 3H) that was eliminated by $CaCl_2$ competition (Fig. 3H).

In silico design of enzyme active sites

We next sought to scaffold the active site of carbonic anhydrase II, which catalyzes the interconversion of carbon dioxide and bicarbonate and has recently been of interest for carbon sequestration (*31*–*33*). The active site consists of three Zn^{2+}-coordinating histidines on two strands and a threonine on a loop, which orients the CO_2 (table S1). Despite the complexity of the irregular, discontinuous three-segment site, hallucination was able to generate designs with sub-angstrom motif AF-RMSDs with correct His placement for Zn^{2+} coordination (Fig. 4A and fig. S9D); these are less than 100 residues in size, considerably smaller than the 261-residue native protein.

We next scaffolded the catalytic side chains of Δ^5-3-ketosteroid isomerase (KSI) (table S1) involved in steroid hormone biosynthesis (*36*). We attempted to use gradient descent by backpropagation through AF (materials and methods; a side chain–predicting version of RF was not available at the time) but found it difficult to obtain accurate side-chain placement; the landscape may be too rugged with the high-resolution side chain–based loss (supplementary text). Better results were obtained with a two-stage approach using, first, both AF and trRosetta (to smoothen the loss landscape) and a description of the active site at the backbone level, followed by a second all-atom AF-only stage once the overall backbone was roughly in place. This yielded multiple plausible solutions with nearly exact matches to the catalytic side-chain geometry (Fig. 4, C and D, and fig. S9E). In silico validation with a held-out AF model (materials and methods) recapitulated the designed active sites. The use of stage-specific loss functions illustrates the ready customizability of the hallucination approach to specific design challenges without network retraining.

Designing protein-binding proteins

To design binders to the cancer checkpoint protein PD-L1, we scaffolded two discontiguous segments of the interfacial β sheet from a high-affinity mutant of PD-1 (Fig. 5A; materials and methods) (*15*). Inpainting yielded designs with not only good AF predictions of the binder monomer (AF pLDDT > 80, motif AF-RMSD < 1.4 Å) but also of the complex between the binder and PD-L1, with an interchain predicted alignment error (inter-PAE) of <10 Å (materials and methods). In contrast to our initial efforts with trRosetta hallucination (fig. S1; supplementary text), it was not necessary to redesign the inpainted sequences using Rosetta. Of 31 designs selected for experimental testing, one design, pdl1_inp_1, bound PD-L1 with a K_d of 326 nM (Fig. 5, B and C), worse than high-affinity consensus (HAC) PD-1 (K_d = 110 pM) (*37*) but better than wild-type PD-1 (K_d = 3.9 μM) (*37*). The pdl1_inp_1 design expressed as a monomer (fig. S15E), was thermostable, and had a CD spectrum consistent with that of a mixed α-β fold (fig. S15F). Unlike native PD-1, which has an immunoglobulin family β-sandwich fold, pdl1_inp_1 has two helices buttressing the interfacial β sheet, as well as an additional fifth inpainted strand extending the interface (fig. S15, A and B). The closest Protein Data Bank (PDB) (*38*) hit had a TM-score of 0.61, and the closest Basic Local Alignment Search Tool (BLAST) NR hit had a sequence identity of 25.4%.

We next used our methods to design ligands engaging multiple receptor binding sites. The nerve growth factor (NGF) receptor TrkA dimerizes upon ligand binding (*39*), and starting from the TrkA-NGF crystal structure, we positioned helical segments derived from two copies of a previously designed TrkA binding protein (*4*) and used hallucination followed by inpainting (materials and methods) to scaffold them on a single chain (Fig. 5, D and E). A design predicted to be well structured (AF pLDDT > 80) and interact with TrkA (inter-PAE < 10 Å) was expressed, purified, and found to bind TrkA, as assessed by biolayer interferometry (BLI) (Fig. 5F). A double mutant that knocked out both designed binding sites abolished TrkA binding, whereas single mutants knocking out either one of the binding sites maintained partial binding (Fig. 5F and fig. S16), suggesting that the protein binds two molecules of TrkA, as designed.

RoseTTAFold is able to predict the structures of protein complexes (*40*), and we hypothesized that it could generate additional binding interactions between hallucinated or inpainted binder and a target beyond the scaffolded motif. We used a "two-chain" hallucination protocol (fig. S17; materials and methods) to design binders to the Mdm2 oncogene by scaffolding the native N-terminal helix of the tumor suppressor protein p53 and obtained diverse designs with AF inter-PAE < 7 Å, target-aligned binder RMSD < 5 Å, binder pLDDT > 85, and spatial aggregation propensity (SAP) score < 35 (fig. S17, D and E); three examples are shown in Fig. 5G.

The above approaches to protein-binder design require starting from a previously known binding motif, but hallucination should in principle be able to generate de novo interfaces as well. To test this, we used two-chain hallucination to optimize 12-residue peptides for binding to 12 targets starting from random sequences, minimizing an interchain entropy loss (fig. S17H). Most of the hallucinated peptides bound at native protein interaction sites (fig. S18A); the remainder bound in hydrophobic grooves resembling protein binding sites (fig. S18B). We used the same procedure to generate 55- to 80-residue binders against TrkA and PDL-1 without starting motif information and obtained designs predicted by AF to complex with the target, at the native ligand binding site, with a target-aligned binder RMSD < 5 Å and an inter-PAE < 10 Å (fig. S17, F and G).

Unlike classical protein design pipelines, which treat backbone generation and sequence design as two separate problems, our methods simultaneously generate both sequence and structure, taking advantage of the ability of RoseTTAFold to reason over and jointly optimize both data types. This results in excellent performance in both generating protein backbones with a geometry capable of hosting a desired site and sequences that strongly encode these backbones. Our hallucinated and inpainted backbones accommodate all of the tested functional sites much more accurately than any naturally occurring protein in the PDB or AF predictions database (fig. S20 and

table S3; supplementary text (*41*), and our designed structures are predicted more confidently from their (single) sequences than most native proteins with known crystal structures and are on par with structurally validated de novo designed proteins (fig. S7, A and B). The hallucination and inpainting approaches are complementary: Hallucination can generate diverse scaffolds for minimalist functional sites but is computationally expensive because it requires a forward and backward pass through the neural network to calculate gradients for each optimization step (materials and methods), whereas inpainting usually requires larger input motifs but is much less compute-intensive and outperforms the hallucination method when more starting information is provided. This difference in performance can be understood by considering the manifold in sequence-structure space corresponding to folded proteins. The inpainting approach can be viewed as projecting an incomplete input sequence-structure pair onto the subset of the manifold of folded proteins (as represented by RoseTTAFold) containing the functional site—if insufficient starting information is provided, this projection is not well determined, but with sufficient information, it produces protein-like solutions, updating sequence and structure information simultaneously. The loss function used in the hallucination approach is constructed with the goal that minima lie in the protein manifold, but there will likely not be a perfect correspondence, and hence stochastic optimization of the loss function in sequence space may not produce solutions that are as protein-like as those from the inpainting approach.

Conclusion

The approaches for scaffolding functional sites presented here require no inputs other than the structure and sequence of the desired functional site and, unlike previous methods, do not require specifying the secondary structure or topology of the scaffold and can simultaneously generate both sequence and structure. Despite a recent surge of interest in using machine learning to design protein sequences (*42*–*49*), the design of protein structure is relatively underexplored, likely because of the difficulty of efficiently representing and learning structure (*50*). Generative adversarial networks and variational autoencoders have been used to generate protein backbones for specific fold families (*51*–*53*), whereas our approach leverages the training of RoseTTAFold on the entire PDB to generate an almost unlimited diversity of new structures and enable the scaffolding of any desired constellation of functional residues. Our "activation maximization" hallucination approach extends related work in this area (*54*–*56*) by leveraging its key strength, the ability to use arbitrary loss functions tailored to specific problems and design any length sequence without retraining. The ability of our inpainting approach to expand from a given functional site to generate a coherent sequence-structure pair should find wide application in protein design because of its speed and generality. The two approaches individually, and the combination of the two, should increase in power as more-accurate protein structure, interface, and small-molecule binding prediction networks are developed.

REFERENCES AND NOTES

1. D. Röthlisberger *et al*., *Nature* **453**, 190–195 (2008).
2. J. B. Siegel *et al*., *Science* **319**, 1387–1391 (2008).
3. J. B. Siegel *et al*., *Science* **329**, 309–313 (2010).
4. L. Cao *et al*., *Nature* **605**, 551–560 (2022).
5. A. Chevalier *et al*., *Nature* **550**, 74–79 (2017).
6. E. Procko *et al*., *Cell* **157**, 1644–1656 (2014).
7. B. E. Correia *et al*., *Nature* **507**, 201–206 (2014).
8. D.-A. Silva *et al*., *Nature* **565**, 186–191 (2019).
9. F. Sesterhenn *et al*., *Science* **368**, eaay5051 (2020).
10. C. Yang *et al*., *Nat. Chem. Biol.* **17**, 492–500 (2021).
11. J. Yang *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 1496–1503 (2020).
12. I. Anishchenko *et al*., *Nature* **600**, 547–552 (2021).
13. C. Norn *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2017228118 (2021).
14. D. Tischer *et al*., bioRxiv 2020.11.29.402743 [Preprint] (2020); https://doi.org/10.1101/2020.11.29.402743.
15. R. Pascolutti *et al*., *Structure* **24**, 1719–1728 (2016).
16. M. Baek *et al*., *Science* **373**, 871–876 (2021).
17. J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, arXiv:1810.04805 [cs.CL] (2019).
18. R. A. Yeh *et al*., arXiv:1607.07539 [cs.CV] (2017).
19. Z. Li, S. P. Nguyen, D. Xu, Y. Shang, *Proc. Int. Conf. Tools. Artif. Intell.* **29**, 1085–1091 (2017).
20. N. Anand, P. Huang, in *Advances in Neural Information Processing Systems 31*, S. Bengio *et al*., Eds. (Curran Associates, Inc., 2018), pp. 7494–7505.
21. J. Jumper *et al*., *Nature* **596**, 583–589 (2021).
22. R. Chowdhury *et al*., bioRxiv 2021.08.02.454840 [Preprint] (2021); https://doi.org/10.1101/2021.08.02.454840.
23. K. T. Simons, R. Bonneau, I. Ruczinski, D. Baker, *Proteins* **37** (suppl. 3), 171–176 (1999).
24. T.-E. Kim *et al*., bioRxiv 2021.12.17.472837 [Preprint] (2021); https://doi.org/10.1101/2021.12.17.472837.
25. M. A. Pak *et al*., bioRxiv 2021.09.19.460937 [Preprint] (2021); https://doi.org/10.1101/2021.09.19.460937.
26. G. R. Buel, K. J. Walters, *Nat. Struct. Mol. Biol.* **29**, 1–2 (2022).
27. J. J. Mousa, N. Kose, P. Matta, P. Gilchuk, J. E. Crowe Jr., *Nat. Microbiol.* **2**, 16271 (2017).
28. T. W. Linsky *et al*., *Science* **370**, 1208–1214 (2020).
29. F. Frolow, A. J. Kalb, J. Yariv, *Nat. Struct. Biol.* **1**, 453–460 (1994).
30. A. Lombardi, F. Pirro, O. Maglio, M. Chino, W. F. DeGrado, *Acc. Chem. Res.* **52**, 1148–1159 (2019).
31. J. R. Calhoun *et al*., *Biopolymers* **80**, 264–278 (2005).
32. A. M. Keech *et al*., *J. Biol. Chem.* **272**, 422–429 (1997).
33. E. N. G. Marsh, W. F. DeGrado, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5150–5154 (2002).
34. M. Yáñez, J. Gil-Longo, M. Campos-Toimil, in *Calcium Signaling*, Md. S. Islam, Ed., vol. 740 of *Advances in Experimental Medicine and Biology* (Springer Netherlands, 2012), pp. 461–482.
35. S. J. Caldwell *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 30362–30369 (2020).
36. H.-S. Cho *et al*., *J. Biol. Chem.* **274**, 32863–32868 (1999).
37. R. L. Maute *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6506–E6514 (2015).
38. H. M. Berman *et al*., *Nucleic Acids Res.* **28**, 235–242 (2000).
39. C. Wiesmann, M. H. Ultsch, S. H. Bass, A. M. de Vos, *Nature* **401**, 184–188 (1999).
40. I. R. Humphreys *et al*., *Science* **374**, eabm4805 (2021).
41. K. Tunyasuvunakool *et al*., *Nature* **596**, 590–596 (2021).
42. J. Ingraham, V. K. Garg, R. Barzilay, T. Jaakkola, "Generative models for graph-based protein design," 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada, 8 to 14 December 2019.
43. A. Strokach, D. Becerra, C. Corbi-Verge, A. Perez-Riba, P. M. Kim, *Cell Syst.* **11**, 402–411.e4 (2020).
44. S. Biswas, G. Khimulya, E. C. Alley, K. M. Esvelt, G. M. Church, *Nat. Methods* **18**, 389–396 (2021).
45. D. Repecka *et al*., *Nat. Mach. Intell.* **3**, 324–333 (2021).
46. J.-E. Shin *et al*., *Nat. Commun.* **12**, 2403 (2021).
47. Z. Wu, K. E. Johnston, F. H. Arnold, K. K. Yang, *Curr. Opin. Chem. Biol.* **65**, 18–27 (2021).
48. N. Anand *et al*., *Nat. Commun.* **13**, 746 (2022).
49. A. Madani *et al*., bioRxiv 2021.07.18.452833 [Preprint] (2021); https://doi.org/10.1101/2021.07.18.452833.
50. S. Ovchinnikov, P.-S. Huang, *Curr. Opin. Chem. Biol.* **65**, 136–144 (2021).
51. N. Anand, R. Eguchi, P.-S. Huang, "Fully differentiable full-atom protein backbone generation," Seventh International Conference on Learning Representations (ICLR 2019), New Orleans, Louisiana, 6 to 9 May 2019.
52. R. R. Eguchi, C. A. Choe, P.-S. Huang, *PLOS Comput. Biol.* **18**, e1010271 (2022).
53. Z. Lin, T. Sercu, Y. LeCun, A. Rives, "Deep generative models create new and diverse protein structures," 35th Conference on Neural Information Processing Systems (NeurIPS 2021), 6 to 14 December 2021.
54. M. Jendrusch, J. O. Korbel, S. K. Sadiq, bioRxiv 2021.10.11.463937 [Preprint] (2021); https://doi.org/10.1101/2021.10.11.463937.
55. L. Moffat, J. G. Greener, D. T. Jones, bioRxiv 2021.08.24.457549 [Preprint] (2021); https://doi.org/10.1101/2021.08.24.457549.
56. L. Moffat, S. M. Kandathil, D. T. Jones, bioRxiv 2022.01.27.478087 [Preprint] (2022); https://doi.org/10.1101/2022.01.27.478087.
57. L. Li *et al*., *J. Phys. Chem. C* **112**, 12219–12224 (2008).
58. J. Wang *et al*., RFDesign: Protein hallucination and inpainting with RosettaFold, version 2, Zenodo (2022); https://doi.org/10.5281/zenodo.6808038.

ACKNOWLEDGMENTS

We thank L. Goldschmidt and K. VanWormer, respectively, for maintaining the computational and wet lab resources at the Institute for Protein Design; C. Norn for general discussions about trRosetta; B. Coventry for advice on interface design; C. Goverde for advice on RSV-F epitopes and motif grafting methods; T. Yu, G. R. Lee, L. An, and X. Wang for advice on flow cytometry; R. Dong and V. Muhunthan for exploratory analyses; N. Hiranuma for exploratory RoseTTAFold training sessions; B. Trippe for feedback on the manuscript; S. Pellock for expertise on enzyme design; A. Fitzgibbon for conceptual discussions on training RoseTTAFold; and C. Garcia for providing biotinylated TrkA. **Funding:** We thank Microsoft for support and for providing Azure computing resources. This work was supported with funds provided by the Audacious Project at the Institute for Protein Design (D.B. and A.S.); a Microsoft gift (M.B. and J.D.); Eric and Wendy Schmidt by recommendation of the Schmidt Futures (D.J.); the DARPA Synergistic Discovery and Design project HR001117S0003 contract FA8750-17-C-0219 (D.B. and W.Y.); the DARPA Harnessing Enzymatic Activity for Lifesaving Remedies project HR001120S0052 contract HR0011-21-2-0012 (N.B.); the Washington Research Foundation (J.W.); the Open Philanthropy Project Improving Protein Design Fund (D.B. and D.T.); Amgen (S.L.); the Human Frontier Science Program Cross Disciplinary Fellowship (LT000395/2020-C) and EMBO Non-Stipendiary Fellowship (ALTF 1047-2019) (L.F.M.); the EMBO Fellowship (ALTF 191-2021) (T.S.); European Molecular Biology Organization Grant (ALTF 139-2018) (B.I.M.W.); the "la Caixa" Foundation (M.E.); the National Institute of Allergy and Infectious Diseases (NIAID) Federal Contract HHSN272201700059C (I.A.), NIH grant DP5OD026389 (S.O.); the National Science Foundation MCB 2032259 (S.O.); the Howard Hughes Medical Institute (D.B., R.R., and K.M.C.), the National Institute on Aging grant 5U19AG065156 (D.B., J.L.W., D.R.H., and M.E.); the National Cancer Institute grant R01CA240339 (D.B. and J.-H.C.); Swiss National Science Foundation (K.M.C. and B.C.); Swiss National Center of Competence for Molecular Systems Engineering (K.M.C. and B.C.); Swiss National Center of Competence in Chemical Biology (K.M.C. and B.C.); and European Research Council grant 716058 (K.M.C. and B.C.). **Author contributions:** Designed the research: J.W., S.L., D.J., D.T., J.L.W., S.O., and D.B. Developed the motif-constrained hallucination method: J.W., D.T., S.L., I.A., and S.O. Contributed code and ideas for hallucination: M.B. and J.D. Generated designs using hallucination: J.W., S.L., D.T., and S.O. Developed the inpainting method: D.J. and J.L.W. Contributed code and ideas for inpainting: M.B., J.W., S.L., and D.T. Generated designs using inpainting: D.J., J.L.W., and A.S. Analyzed data: J.W., S.L., D.J., D.T., J.L.W., and M.E. Trained neural networks: D.J., J.L.W., and M.B. Performed RSV-F experiments: K.M.C., R.R., L.F.M., and J.W. Performed di-iron experiments: J.L.W. and D.J. Performed EF-hand experiments: A.S. and J.L.W.

Performed PD-L1 experiments: W.Y., D.R.H., J.W., S.L., and D.J. Contributed reagents and technical expertise: T.S., J.-H.C., L.F.M., N.B., B.I.M.W., B.C., A.M., and F.D. Wrote the manuscript: J.W., D.J., J.L.W., S.L., D.T., S.O., and D.B. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Code and neural network weights are available at https://github.com/RosettaCommons/RFDesign and https://github.com/sokrypton/ColabDesign and archived at Zenodo (*58*). Plasmids of designed proteins are available upon request.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn2100
Materials and Methods
Supplementary Text
Figs. S1 to S21
Tables S1 to S5
Algorithm S1
References (*59*–*86*)
MDAR Reproducibility Checklist
Data S1 and S2

Submitted 11 November 2021; accepted 24 June 2022
10.1126/science.abn2100

SURFACE CHEMISTRY

Quantum effects in thermal reaction rates at metal surfaces

Dmitriy Borodin[1,2]*, Nils Hertl[1,2], G. Barratt Park[1,2,3], Michael Schwarzer[1], Jan Fingerhut[1], Yingqi Wang[4], Junxiang Zuo[4], Florian Nitz[1], Georgios Skoulatakis[2], Alexander Kandratsenka[2], Daniel J. Auerbach[2], Dirk Schwarzer[2], Hua Guo[4], Theofanis N. Kitsopoulos[1,2,5,6], Alec M. Wodtke[1,2]*

There is wide interest in developing accurate theories for predicting rates of chemical reactions that occur at metal surfaces, especially for applications in industrial catalysis. Conventional methods contain many approximations that lack experimental validation. In practice, there are few reactions where sufficiently accurate experimental data exist to even allow meaningful comparisons to theory. Here, we present experimentally derived thermal rate constants for hydrogen atom recombination on platinum single-crystal surfaces, which are accurate enough to test established theoretical approximations. A quantum rate model is also presented, making possible a direct evaluation of the accuracy of commonly used approximations to adsorbate entropy. We find that neglecting the wave nature of adsorbed hydrogen atoms and their electronic spin degeneracy leads to a 10× to 1000× overestimation of the rate constant for temperatures relevant to heterogeneous catalysis. These quantum effects are also found to be important for nanoparticle catalysts.

Enormous effort has gone into developing predictive theories of thermal reaction rates (*1*), with one goal being accurate kinetic models of heterogeneous catalysis, an industrial cornerstone of modern society (*2*). Modeling real catalytic reactors presents technical problems because they often involve networks of reactions (*3*, *4*), complicating meaningful comparisons to experiment that could test a theory's assumptions. A possible solution is to compare experiment and theory using simplified model systems that involve only a single elementary reaction. Unfortunately, even this comparison is seldom achieved because accurate measurements of elementary reaction rates are rare in surface chemistry (*5*).

Illustrative of these problems is the thermal recombination of H atoms on transition metals, leading to H_2 formation. Being perhaps the simplest reaction for theoretical modeling and omnipresent as an elementary step in industrial catalysis [e.g., hydrogenation of unsaturated fats (*6*), ammonia synthesis (*7*), and electrochemical hydrogen production (*8*)], it is an obvious starting point for the development of accurate rate theories in surface chemistry. Unfortunately, large uncertainties in the experimentally derived second-order rate constants arise because of difficulties in obtaining accurate initial concentrations (*9*). If these and other experimental problems could be overcome, this reaction would provide an ideal system for benchmarking rate theory, especially for testing approximate treatments of quantum effects.

From the study of gas-phase reactions, exact treatments of nuclear quantum effects are often considered to be unnecessary above ~500 K (*10*), and, because most catalytic reactors operate at increased temperatures, one might conclude that a classical approximation (*11*) or approximate ad hoc quantum treatments, like harmonic transition-state theory (hTST) (*12*, *13*), would be sufficient to model surface chemistry. But the need to go beyond hTST has been pointed out recently (*14*) and new methods were reported, although they also lack validation from experiment. Electron spin is another important quantum effect on surface reactions; for example, in H atom recombination, only one out of four electron spin combinations yields a stable H_2 molecule. However, the spin-degeneracy of reactants and products has, to our knowledge, never been included in calculations of reaction rates at metal surfaces.

This paper reports kinetic data for H atom recombination on both the Pt(111) and Pt(332) surfaces obtained with velocity-resolved kinetics (VRK), which was previously used only to study first-order and pseudo–first-order reactions on model catalysts (*15*, *16*). For this work, we have extended VRK to the measurement of rate constants for second-order reactions by measuring the absolute reactant flux, which, when combined with known sticking probabilities (*17*, *18*), provides accurate initial concentrations $[H]_0$ and eliminates the main source of error found in previous work.

To understand the kinetics more deeply, we also constructed a quantum rate model (QRM) that accurately reproduced experimental rate constants over 12 orders of magnitude for temperatures between 250 and 1000 K with no adjustable parameters. Comparison to a corresponding classical rate model (CRM) revealed how large and crucially important quantum effects are; the classical reaction rate constants were ~20 times larger than quantum rate constants even at 1000 K, with an increasing deviation at lower temperatures. For reactions at stepped surfaces, the errors were even higher. This dramatic quantum reduction of the reaction rate resulted from both the delocalization of the adsorbed H* nuclei as well as the influence of electron spin degeneracy.

Results

The experiments are described in detail in the supplementary materials (SM). Briefly, a pulsed molecular beam with a controlled mixture of H_2 and D_2 illuminated either a Pt(111) or Pt(332) crystal facet, with step densities of 0.1 to 0.6% and 16.7%, respectively. The transient rates of HD formation were then recorded using VRK, where pulsed laser-ionization, time-of-flight mass spectrometry reports the product's mass-to-charge ratio (m/Z) and its density as a function of delay between the pulsed molecular and laser beams. Because the ions were detected with slice imaging (*19*, *20*) yielding product velocity, we could

[1]Institute for Physical Chemistry, University of Göttingen, Tammannstraße 6, 37077 Göttingen, Germany. [2]Department of Dynamics at Surfaces, Max Planck Institute for Multidisciplinary Sciences, am Faßberg 11, 37077 Göttingen, Germany. [3]Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, TX 79409-1061, USA. [4]Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, NM 87131, USA. [5]Department of Chemistry, University of Crete, 71003 Heraklion, Greece. [6]Institute of Electronic Structure and Laser, FORTH, 71110 Heraklion, Greece.
***Corresponding author. Email: dborodi@gwdg.de (D.B.); alec.wodtke@mpinat.mpg.de (A.M.W.)**

accurately compute the transient product flux as a function of reaction time at the surface. Initial reactant concentrations are needed to obtain second-order rate constants (SM section S2a). These values were obtained from the absolute flux profiles of the incident molecular beams (SM section S2b) and known sticking coefficients (*17*, *18*) (SM section S3). Finally, VRK data obtained at $m/Z = 2$, 3, and 4 led to isotopic branching fractions (SM section S4), from which we obtained isotope-specific rate constants.

The QRM developed in this work is an exact formulation of a thermal rate constant. It yields accurate isotope-specific thermal rates as long as one has accurate isotope-specific thermal sticking probabilities $\left\langle S_0^{H_2,HD,D_2} \right\rangle(T)$, adsorption energies $E_0^{H_2,HD,D_2}$, and reactant Q_{H^*,D^*} and product Q_{H_2,HD,D_2} partition functions:

$$k_{H_2}(T) = \left\langle S_0^{H_2} \right\rangle(T) \sqrt{\frac{k_B T}{2\pi m_{H_2}}} \frac{Q_{H_2}/V}{(Q_{H^*}/A)^2} \exp\left(-\frac{E_0^{H_2}}{k_B T}\right) \quad (1)$$

where T is temperature, k_B is the Boltzmann constant, V is reference volume, and A is reference area for the partition function evaluation.

The QRM rate constant for H_2 desorption by the recombination of two adsorbed H atoms, $k_{H_2}(T)$, given by Eq. 1, is derived from the principle of detailed balance provided in SM section S5a. Expressions for $k_{HD}(T)$ and $k_{D_2}(T)$ were easily obtained by analogy. Accurate values for $E_0^{H_2,HD,D_2}$ and $\left\langle S_0^{H_2,HD,D_2} \right\rangle(T)$ can be obtained from prior experiments (SM sections S3 and S6). These measured quantities allowed us to avoid errors associated with the theoretical determination of the thermal dissociative adsorption rates and density functional theory (DFT) calculations of adsorption energies, which can be highly dependent on the choice of exchange-correlation functional (*21*). In addition, the partition function for the hydrogen molecule in the gas phase Q_{H_2} is well known.

The adsorbate partition functions Q_{H^*,D^*} are crucial inputs to the QRM and were computed with a quantum potential energy sampling (QPES) method, where the nuclear part of the partition function is obtained by a direct state count. States and energies were obtained by solving the nuclear Schrödinger equation with DFT interaction potentials computed with two different functionals and assuming a static Pt surface (SM section S1b). This procedure was performed for H interacting with both Pt(111) and Pt(332). We found that Q_{H^*} is weakly dependent on the choice of DFT functional (SM section S5e). The electronic contribution to Q_{H^*}, which accounts for the twofold spin degeneracy of the H-Pt system, was explicitly included.

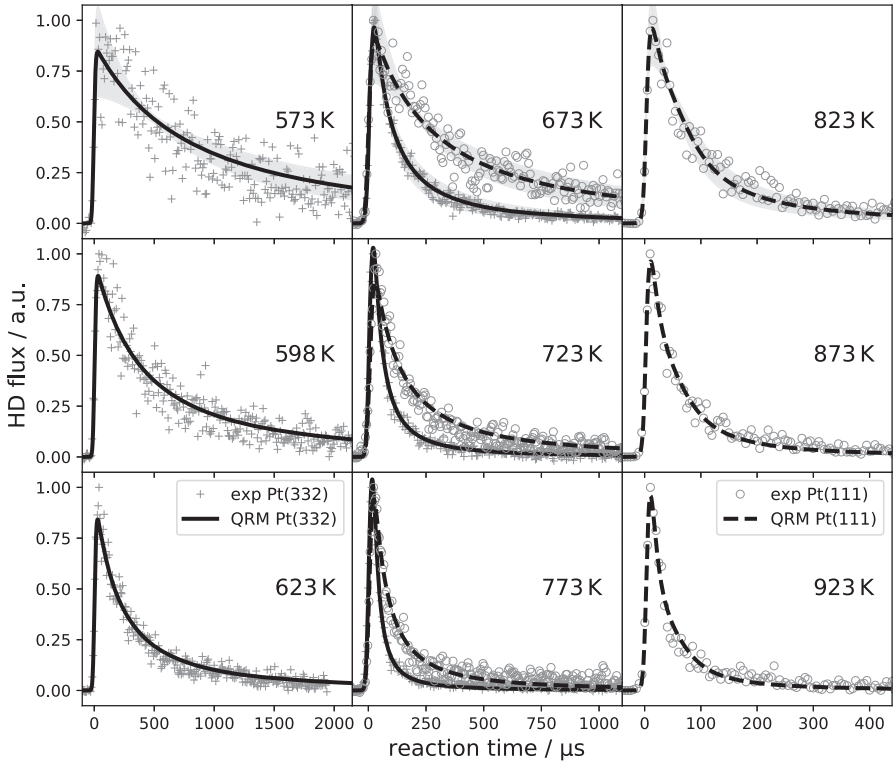


Fig. 1. VRK of H atom recombination on Pt(111) and Pt(332). Measured HD formation rates for Pt(111) (○) and Pt(332) (+) are compared with the results of the QRM (dashed and solid lines). The temperature dependence and the transient rate of the measurements are quantitatively captured by the model for both facets. The shaded regions of the top three panels indicate 2σ uncertainty, mainly associated with the absolute reactant flux measurement (~30%) and the dissociative adsorption energies. The excellent agreement between VRK and QRM is achieved without adjustable parameters. a.u., arbitrary units.

Figure 1 presents the experimentally obtained HD formation rates for reactions on Pt(111) and Pt(332) and compares them with a simulation of the experiment. The simulations used rate constants from the QRM for all three isotopologs (fig. S9) and accounted for the temporal profile of the dosing pulse and its spatial inhomogeneity, $f(t, r)$ (Fig. 2A), where t is time and r is radial distance, as well as reactant diffusion. This aspect of the data analysis goes beyond past work and is essential because the rates of second-order reactions are sensitive to surface concentration distributions and gradients. The full diffusion-reaction model is described in SM section S7 and accounts for well-known diffusion effects on surface reaction rates discussed in previous work (*22*). Figure 2B shows that the isotope effect at these temperatures is small and well described by the QRM. Inspection of Figs. 1 and 2B clearly shows that the QRM, which has no adjustable parameters, reproduces experimental data for reactions on both Pt(111) and Pt(332).

Figure 3 shows the VRK-derived H* recombination rate constants (black circles) compared with those of previous work (light red trapezoids) for reactions on Pt(111). Previous studies used temperature programmed desorption (TPD) for $T < 400$ K (*23*–*26*) and molecular beam relaxation spectrometry (MBRS) for $T > 400$ K (*27*, *28*). The uncertainty in the previously reported rate constants spans three orders of magnitude. We note that previous work studied different isotopic recombination reactions (fig. S8); however, given the small isotope effect found in the present study (Fig. 2B and fig. S9), these differences between experiments cannot explain the large range of reported values. The VRK measurements clearly distinguish the accuracy of two previous MBRS measurements that fall within the uncertainty range of (*27*) but differ by two orders of magnitude from rate constants reported in (*28*).

A hallmark of a fundamentally correct model is its ability to reproduce accurate experimental data over a broad temperature range. The QRM uses a fundamentally correct ab initio adsorbate partition function that leads to excellent agreement with experiment over a large temperature range. The performance of the QRM for Pt(111) at temperatures between 650 and 950 K is demonstrated by comparison to VRK-derived rate constants, whereas low-temperature comparisons rely on TPD. Uncertainties in the TPD-derived rate constants

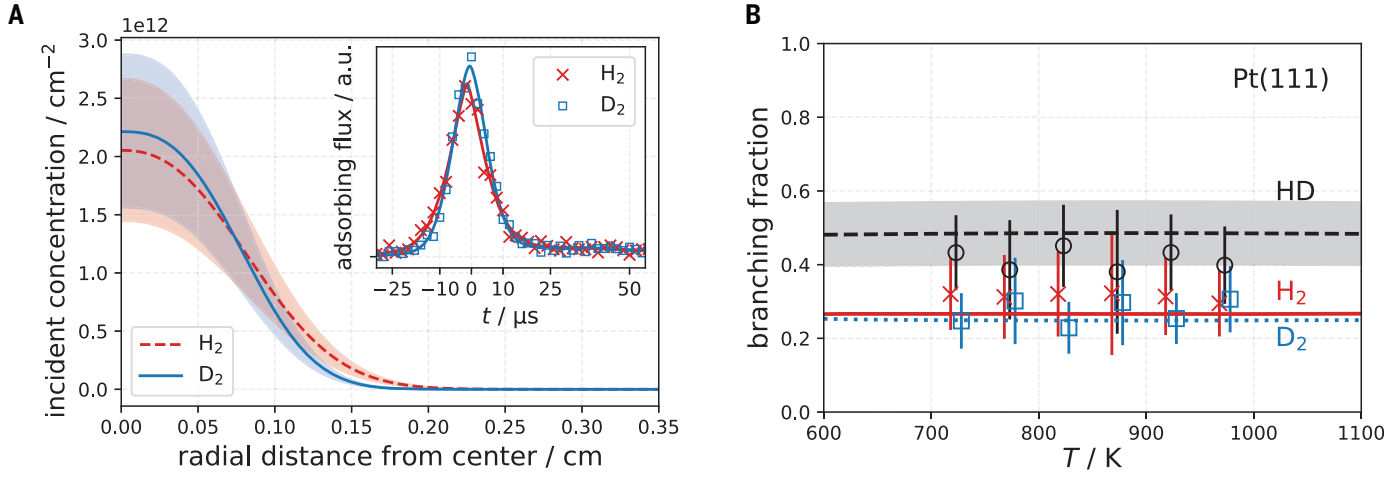


Fig. 2. Calibration of the molecular beam and isotopic branching. (**A**) The space-dependent H_2 and D_2 dosing profiles used to determine the absolute initial concentration of H* and D*. These results were obtained from laser-based calibration of the molecular beam flux and are required to accurately determine the recombination rate constants. The shaded regions indicate the 2σ uncertainty range. The inset shows the temporal profile of the molecular beam pulse. (**B**) Isotopic branching fraction from VRK experiments (symbols) and QRM (lines). The agreement shows that QRM correctly predicts the isotope effect. The error bars and the gray-shaded region reflect 2σ uncertainty in the experiment and model, respectively. Note that some symbols have been shifted by ±5 K for clarity.

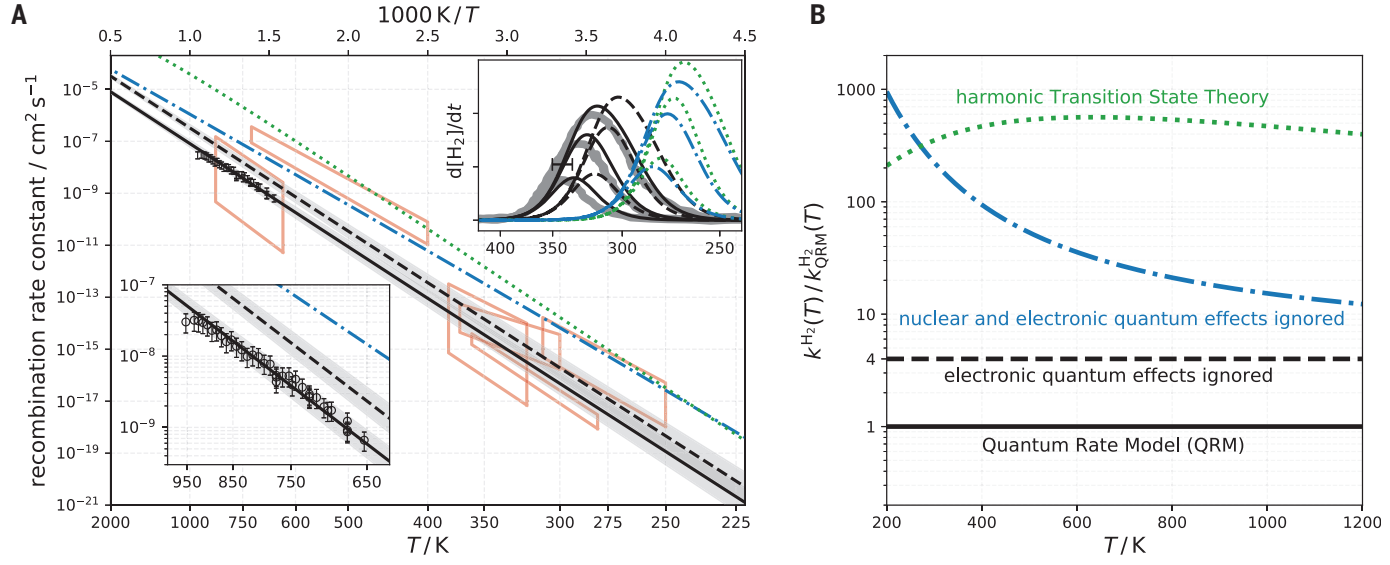


Fig. 3. Rate constants for H atom recombination on Pt(111). (**A**) Light red trapezoids show the temperature-range and rate-constant uncertainties of previous work (*23*–*28*). Shown are experimental results from this work (○) with 2σ error bars compared with the results of the QRM (black solid line), hTST (green dotted line), CRM (blue dash-dotted line), and QRM neglecting electron spin (black dashed line). The inset at the bottom left shows an expanded view. The inset at the top right compares TPD spectra (broad gray lines) from (*29*) with the predictions of the QRM model, QRM neglecting spin, the CPES model, and the hTST model for three initial H* coverages of 0.1, 0.2, and 0.3 ML. The gray-shaded region and the horizontal error bar on one of the modeled TPD spectra reflects the uncertainty of the experimental H_2 chemisorption energy. The ability of the QRM rate constants to quantitatively reproduce experimental data demonstrates the importance of both nuclear and electronic quantum effects. (**B**) Comparison of the approximate predictions of three rate models to QRM rate constants. Neglecting spin degeneracy, using a fully classical approximation or a commonly adopted approximate quantum model both introduce large errors even at high temperatures. Similar errors are seen for recombination rates on the stepped Pt(332) surface (see fig. S14). See fig. S12 for a detailed decomposition of the errors observed from hTST and adsorbate entropy approximations.

arise from questionable approximations used to derive rate constants from the data, neglect of the coverage dependence of adsorption energies (*26*), dubious estimations of prefactors (*25*), and neglect of the influence of steps (*23*). To make the most meaningful comparison, we used the QRM to directly simulate TPD spectra from (*29*), where the influence of steps was carefully identified and removed (SM section S8). Here, we also accounted for the previously reported coverage dependence of the adsorption energy (*24*, *30*) (fig. S11 and SM section S6). The comparison is shown in the top-right inset of Fig. 3A. The solid black lines of the QRM are in excellent agreement with the TPD spectra [broad gray lines, from (*29*)] for three initial coverages.

Discussion

The aforementioned comparisons to kinetics experiments carried out between 250 and 950 K

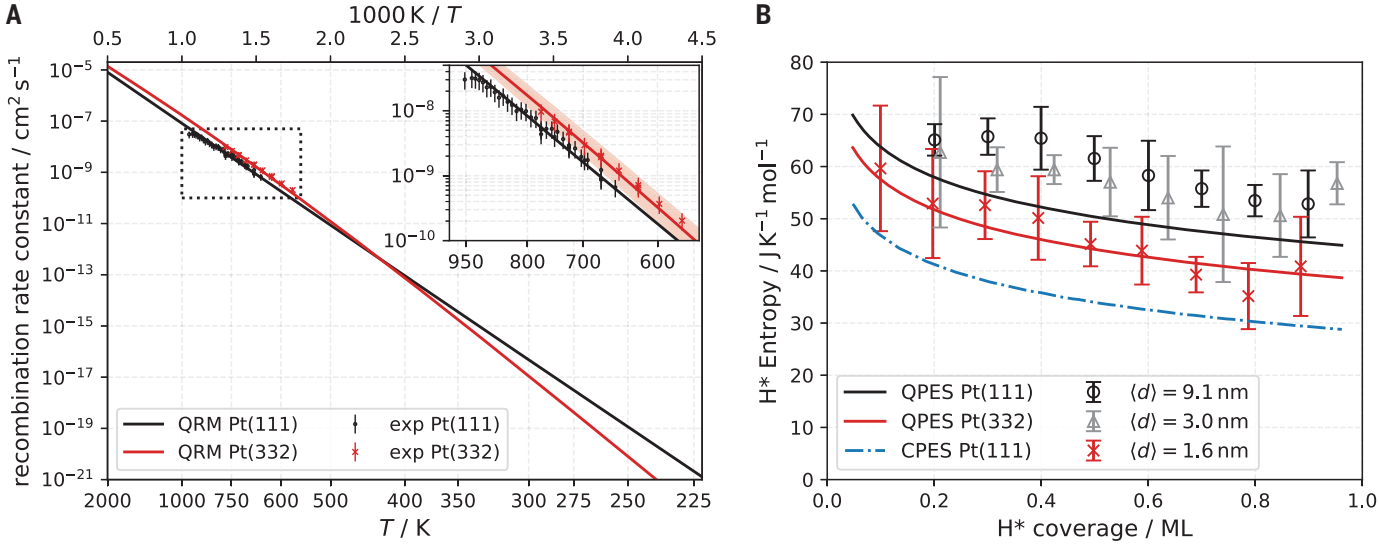


Fig. 4. The influence of steps on H atom recombination on Pt. (**A**) Rate constants derived from VRK experiments (symbols) for H atom recombination on Pt(111) and Pt(332) are compared with QRM predictions (solid lines). The 2σ uncertainty of the rate constants of Pt(332) QRM is shown as a red-shaded region. The inset is a magnification of the area enclosed by the dotted rectangle. (**B**) Entropies obtained experimentally at 598 K (symbols with 2σ error bars) and from CPES (blue dash-dotted line) for H* bound to Pt nanoparticles from (*11*). Also shown are QPES entropies for H* bound to Pt(111) (solid black line) and Pt(332) (solid red line) that were obtained in this work (see SM section S9). The nuclear quantum effect contribution is 12 J mol^{-1} K^{-1}, and the contribution of electron spin is 6 J mol^{-1} K^{-1}. The comparison suggests that the nanoparticle-size dependence of the H* entropy is determined by the concentration of steps.

demonstrated the validity of the QRM rate constants over 12 orders of magnitude and for H atom coverages up to 0.3 monolayer (ML). Within the context of the principle of detailed balance as implemented in the QRM, this coherent picture demonstrates the quantitative consistency of previously reported sticking coefficients and binding energies with the kinetics measurements of this work. The agreement over such a wide range of rates provides confidence in the QRM rate constants, making H recombination on Pt(111) a reliable benchmark for approximate rate theories in surface chemistry.

It is worth noting that the QRM as implemented in this work is semiempirical because it relies on experimental values of thermal sticking coefficients and adsorption energies. However, it also provides a path to an ab initio theory of thermal reaction rates, if these quantities can be accurately calculated from first principles.

The framework of the QRM allows us to critically test the quality of predictions based on approximations that are commonly used in kinetic modeling of heterogeneous catalysis. The results of this analysis are shown in Fig. 3. The most widely used model for rate constants (hTST) introduces quantum effects in an approximate way, where nuclear partition functions are computed assuming separable motion of contributing degrees of freedom that can each be approximated as a harmonic oscillator. By definition, recrossing corrections are not included in hTST (*12*, *13*). The hTST rate constants, calculated by placing the dividing surface far above the surface, are shown as a green dotted line in Fig. 3. This approximation overestimates the experimental reaction rate constant by two to three orders of magnitude at all temperatures between 200 and 1200 K. The major source of errors in hTST arise from the harmonic simplifications made to the H-Pt interaction potential (resulting in errors of a factor 5 to 25) and the neglect of recrossing corrections to TST (with errors of a factor 5 to 10) (see fig. S12A for details).

The next, more sophisticated level of rate theory uses the complete potential energy sampling (CPES) method to characterize entropy associated with the in-plane degrees of freedom of H*. CPES is considered by many to provide the most accurate adsorbate partition function (*31*), and it has been applied to characterize H interaction at metals (*11*). It accounts for anharmonicity by using a semiclassical partition function computed from the adsorbate potential energy surface, which may be obtained with DFT (*11*, *14*, *31*). To evaluate this approach, we modified the QRM, replacing the QPES by the CPES adsorbate partition function but retaining the other parameters in Eq. 1. This substitution serves to illustrate the classical counterpart of the QRM, which we hereafter denote as the CRM. The rate constants predicted by the CRM are shown as blue dash-dotted lines in Fig. 3. The CRM performed better than hTST but nevertheless overestimated the rate constant by a factor of 20, even at temperatures as high as 1000 K. The error is more than 100-fold at 300 K, a temperature typical for electrochemical applications. Although our detailed analysis is focused on Pt(111), the errors introduced by hTST and the CRM are similar for reactions on Pt(332) (fig. S14).

A major source of error in the CPES method arises from the classical description of the adsorbate's in-plane motion. This can be understood by considering that the in-plane zero-point energy of H* on Pt(111) (58 meV) is almost equal to the classical diffusion barrier (60 meV) (see fig. S13). Thus, classical and quantum descriptions of H* motion on the surface lead to very different results. CPES excludes H* from classically forbidden regions of space, whereas quantum mechanically, there is a substantial probability to populate these regions. Furthermore, CPES does not account for the uncertainty principle, which prevents localization of H* at the classical energy minimum at low temperature. The surface area explored by the H atom is underestimated by CPES and thus so too is the adsorbate entropy. This results in an overestimate of the corresponding rate constant. Our results underscore the importance of quantum delocalization and help explain why the deviations of CRM become more severe at low temperatures. Quantum delocalization is also the reason why hTST fails. All quantum states above the ground state exhibit probability maxima at positions far from the potential energy minimum (fig. S13).

We may investigate other sources of error in the CRM by using the QPES partition function but neglecting electron spin. Figure 3A shows rate constants predicted on this basis, and in Fig. 3B, one can see that neglect of electron spin degeneracy led to a 4× overestimate of the rate constant at all temperatures. This

result can be understood intuitively if we consider that when two H* atoms attempt to react, they must approach one another in one of four degenerate states with either parallel (triplet) or antiparallel (singlet) spins. Only the singlet state correlates with the formation of gas-phase singlet H_2 products; hence, including spin degeneracy reduces the reaction rate by a factor of four. This result should not come as a surprise to those familiar with rate calculations for gas-phase reactions where spin degeneracy is routinely included (*1*). Nonetheless, to the best of our knowledge, this is the first demonstration that the rates of thermal reactions at metal surfaces depend on adsorbate spin.

The QRM developed in this work also describes the reaction rate on stepped surfaces. In Fig. 4A, we compare the predicted rate constants of the QRM for Pt(111) and Pt(332) surfaces to those derived from VRK experiments. For details of the QRM treatment of the reaction on Pt(332), see SM sections S1b and S6. Experiment shows that near 700 K, the rate constants for reaction on the (332) facet is larger than that on the (111) facet, an effect that is quantitatively captured by the QRM. This result may appear surprising, because the H atom's binding energy is larger at steps than at terraces (*30*, *32*). A naïve view of Eq. 1 suggests that this leads to a lower rate constant. However, careful analysis of the thermally populated quantum states used in the QPES partition functions showed that at these temperatures, H atoms on the (332) facet tend to remain localized near step sites (fig. S15). This fact reduces their in-plane translational entropy and leads to an increase in the rate constant because the effect of entropy is larger than that produced by a larger step binding energy.

This observation reflects how changing temperature alters the relative influence of energy and entropy on the rate constant. In past work, similarities in TPD spectra of H_2 desorbing from Pt(111) and a B-type stepped Pt surface at $T \sim 350$ K were taken as evidence for a lack of preferential step binding (*26*). Inspection of QRM rate constants in Fig. 4A reveals that at 350 K, the similarity in desorption rate constants arises from compensation between energetic and entropic contributions (see SM section S8 for details). Our work supports conclusions derived from He-atom and ion-scattering experiments that H binds more strongly to B-type steps (*30*, *32*). Only at much lower temperatures does the energetic preference for H binding at steps cause the rate constant on the stepped surface to drop below that on (111) terraces.

Because the QRM approach developed in this work provides an accurate determination of rate constants on stepped surfaces, we expect that QPES entropies used in the QRM would help us understand experiments performed on size-selected Pt nanoparticles (*11*), because smaller nanoparticles exhibit higher step concentrations. Figure 4B shows measured H* entropies for Pt nanoparticles of various sizes reproduced from (*11*)—the entropy increases with the nanoparticle size, consistent with observations presented above that Pt steps reduce H* entropy. Also shown are CPES entropies reported in (*11*), which fail to describe entropies derived from experiment. Notably, the entropies found using the QPES method for H* bound to Pt(111) (see Fig. 4B) are in good agreement with entropies for the largest particle sizes. Note that surfaces of large nanoparticles are primarily composed of the (111) facets (*11*). Figure 4B also shows QPES entropies for H* on the (332) facet, which compare well to experimentally obtained entropies on small nanoparticles. This comparison further supports our hypothesis that the H* entropy decreases as nanoparticle size decreases and the relative importance of step defects increases. This result points out the importance of quantum effects even for the description of thermodynamic state functions in advanced catalytic materials.

Conclusion

As this work has shown, H* recombination on Pt surfaces exhibits large quantum effects even at increased temperatures relevant to catalysis. These quantum effects in the reaction rates and in the thermodynamic properties of the adsorbed H atoms arise in part from the H atom's light mass, where a careful treatment of its wave properties is required to obtain accurate results. Such nuclear quantum effects will diminish in importance for heavier adsorbates. However, the effect of spin degeneracy demonstrated here will remain of general importance for a host of reactions of heavier species involved in real-world catalysis. At present, it is not easily possible to determine the lowest-energy spin state for metal surfaces with DFT. Developing theoretical and experimental methods that are able to probe the general influence of spin on reaction rates presents the next challenge on the way toward fully predictive surface chemistry at metal catalysts.

REFERENCES AND NOTES

1. J. L. Bao, D. G. Truhlar, *Chem. Soc. Rev.* **46**, 7548–7596 (2017).
2. F. Studt, *Front. Catal.* **1**, 658965 (2021).
3. A. H. Motagamwala, J. A. Dumesic, *Chem. Rev.* **121**, 1049–1076 (2021).
4. J. E. Sutton, W. Guo, M. A. Katsoulakis, D. G. Vlachos, *Nat. Chem.* **8**, 331–337 (2016).
5. G. B. Park *et al*., *Nat. Rev. Chem.* **3**, 723–732 (2019).
6. P. Sabatier, *Catalysis in Organic Chemistry* (D. Van Nostrand Company, 1922).
7. G. Ertl, *Angew. Chem. Int. Ed.* **47**, 3524–3535 (2008).
8. A. Godula-Jopek, *Hydrogen Production by Electrolysis* (Wiley-VCH, 2015).
9. M. P. D'Evelyn, R. J. Madix, *Surf. Sci. Rep.* **3**, 413–495 (1983).
10. Y. V. Suleimanov, F. J. Aoiz, H. Guo, *J. Phys. Chem. A* **120**, 8488–8502 (2016).
11. M. García-Diéguez, D. D. Hibbitts, E. Iglesia, *J. Phys. Chem. C* **123**, 8447–8462 (2019).
12. S. Bhandari, S. Rangarajan, C. T. Maravelias, J. A. Dumesic, M. Mavrikakis, *ACS Catal.* **10**, 4112–4126 (2020).
13. M. Jørgensen, H. Grönbeck, *ACS Catal.* **6**, 6730–6738 (2016).
14. A. Bajpai, P. Mehta, K. Frey, A. M. Lehmer, W. F. Schneider, *ACS Catal.* **8**, 1945–1954 (2018).
15. D. Borodin *et al*., *J. Am. Chem. Soc.* **143**, 18305–18316 (2021).
16. J. Neugebohren *et al*., *Nature* **558**, 280–283 (2018).
17. R. van Lent *et al*., *Science* **363**, 155–157 (2019).
18. A. C. Luntz, J. K. Brown, M. D. Williams, *J. Chem. Phys.* **93**, 5240–5246 (1990).
19. K. Tonokura, T. Suzuki, *Chem. Phys. Lett.* **224**, 1–6 (1994).
20. C. R. Gebhardt, T. P. Rakitzis, P. C. Samartzis, V. Ladopoulos, T. N. Kitsopoulos, *Rev. Sci. Instrum.* **72**, 3848–3853 (2001).
21. G. J. Kroes, *Phys. Chem. Chem. Phys.* **23**, 8962–9048 (2021).
22. D. R. Olander, *J. Colloid Interface Sci.* **58**, 169–183 (1977).
23. K. Christmann, G. Ertl, T. Pignet, *Surf. Sci.* **54**, 365–392 (1976).
24. B. Poelsema, K. Lenz, G. Comsa, *J. Phys. Condens. Matter* **22**, 304006 (2010).
25. P. Samson, A. Nesbitt, B. E. Koel, A. Hodgson, *J. Chem. Phys.* **109**, 3255–3264 (1998).
26. M. J. van der Niet, A. den Dunnen, L. B. Juurlink, M. T. Koper, *J. Chem. Phys.* **132**, 174705 (2010).
27. G. E. Gdowski, J. A. Fair, R. J. Madix, *Surf. Sci.* **127**, 541–554 (1983).
28. M. Salmerón, R. J. Gale, G. A. Somorjai, *J. Chem. Phys.* **70**, 2807–2818 (1979).
29. S. K. Jo, *Surf. Sci.* **635**, 99–107 (2015).
30. B. J. J. Koeleman, S. T. de Zwart, A. L. Boers, B. Poelsema, L. K. Verhey, *Nucl. Instrum. Methods Phys. Res.* **218**, 225–229 (1983).
31. M. Jørgensen, H. Grönbeck, *J. Phys. Chem. C* **121**, 7199–7207 (2017).
32. B. Poelsema, G. Mechtersheimer, G. Comsa, *Surf. Sci.* **111**, 519–544 (1981).
33. D. Borodin, Hydrogen atom recombination on Pt(111) and Pt(332). Zenodo (2022); doi:10.5281/zenodo.6565004.

ACKNOWLEDGMENTS

We thank J. C. Tully for helpful discussions. **Funding:** D.B. and M.S. thank the BENCh graduate school, funded by the DFG (389479699/GRK2455). T.N.K., G.S., A.K., M.S., and J.F. acknowledge support from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement no. 833404). Y.W., J.Z., and H.G. acknowledge the US National Science Foundation (grant. no. CHE-1951328), and H.G. thanks the Alexander von Humboldt Foundation for a Humboldt Research Award. The calculations were partially performed at the Center for Advanced Research Computing (CARC) at the University of New Mexico and at the National Energy Research Scientific Computing (NERSC) Center. **Author contributions:** D.B., M.S., and J.F. conducted the transient kinetics experiments. Flux calibration procedures were developed by D.B., G.B.P., M.S., F.N., D.J.A., and T.N.K. D.B. and D.S. developed the reaction-diffusion analysis. D.B. and A.M.W. developed the quantum rate model. N.H., Y.W., J.Z., and H.G. conducted DFT calculations. D.B., N.H., A.K., and H.G. analyzed DFT calculations and developed methods for description of nuclear partition functions. M.S., J.F., G.S., T.N.K., D.J.A., D.S., and A.K. participated in discussion of the results. D.B., D.J.A., H.G., and A.M.W. wrote the manuscript and the supporting material. All authors contributed to the reviews of the manuscript and the supporting material. **Competing interests:** None declared. **Data and material availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials and are publicly available in the Zenodo repository (*33*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq1414
Materials and Methods
Supplementary Text
Figs. S1 to S15
References (*34*–*71*)

Submitted 21 March 2022; accepted 16 June 2022
10.1126/science.abq1414

EVOLUTION

A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes

Emily R. Hager[1]†‡, Olivia S. Harringmeyer[1]†, T. Brock Wooldridge[1], Shunn Theingi[1], Jacob T. Gable[1], Sade McFadden[1], Beverly Neugeboren[1], Kyle M. Turner[1]§, Jeffrey D. Jensen[2], Hopi E. Hoekstra[1]*

How locally adapted ecotypes are established and maintained within a species is a long-standing question in evolutionary biology. Using forest and prairie ecotypes of deer mice (*Peromyscus maniculatus*), we characterized the genetic basis of variation in two defining traits—tail length and coat color—and discovered a 41-megabase chromosomal inversion linked to both. The inversion frequency is 90% in the dark, long-tailed forest ecotype; decreases across a habitat transition; and is absent from the light, short-tailed prairie ecotype. We implicate divergent selection in maintaining the inversion at frequencies observed in the wild, despite high levels of gene flow, and explore fitness benefits that arise from suppressed recombination within the inversion. We uncover a key role for a large, previously uncharacterized inversion in the evolution and maintenance of classic mammalian ecotypes.

Wide-ranging species that occupy diverse habitats often evolve distinct ecotypes—intraspecific forms that differ in heritable traits relevant to their local environments (*1*). Ecotypes frequently differ in multiple locally adaptive phenotypes (*2*), and although ecotypes sometimes show partial reproductive isolation (*2*), many experience substantial intraspecific gene flow (*3*). This raises an important question: How are differences in multiple traits maintained between ecotypes when migration acts as a homogenizing force?

One explanation is that natural selection keeps each locus associated with locally adaptive trait variation at migration-selection equilibrium (*4*). However, in cases of high migration, this requires strong selection acting on many independent alleles. Linkage disequilibrium can play an important role by allowing linked loci, each with potentially weaker selective effects, to establish and be maintained together (*5*), which can lead to concentrated genetic architectures of ecotype-specific traits (*6*). Characterizing the genetic basis of the full set of ecotypic differences and the role of migration, selection, and recombination in maintaining these differences is thus critical to understanding local adaptation specifically and biological diversification more generally.

[1]Department of Molecular and Cellular Biology, Department of Organismic and Evolutionary Biology, Museum of Comparative Zoology, and Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138, USA. [2]School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA.
*Corresponding author. Email: hoekstra@oeb.harvard.edu
†These authors contributed equally to this work. ‡Present address: Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA. §Present address: Centre for Teaching Support & Innovation, University of Toronto, Toronto, ON M5S 3H1, Canada.

One of the most abundant and widespread mammals in North America is the deer mouse (*Peromyscus maniculatus*), which is continuously distributed across diverse habitats from the Arctic Circle to central Mexico. In the early 1900s, a taxonomic revision of this species described two distinct ecotypes: a forest and a prairie form (*7*). Several features distinguish the semiarboreal forest mice that occupy dark-soil habitats from their more terrestrial prairie counterparts that occupy light substrates. Most notably, forest mice typically have longer tails and darker coats than those of prairie mice (*7*–*9*), with large differences in these traits maintained between ecotypes despite evidence for gene flow (*10*, *11*). This consistent divergence in multiple traits provides an opportunity to test the mechanisms that establish and maintain ecotypes.

Forest and prairie mice differ in multiple traits

To study divergence between the forest and prairie ecotypes, we selected two focal populations—one from a coastal temperate rainforest (*P. m. rubidus*, referred to hereafter as the forest ecotype) and one from an arid sagebrush steppe habitat (*P. m. gambelii*, referred to as the prairie ecotype) in the northwestern US—separated by ~500 km (Fig. 1A). After establishing laboratory colonies from wild-caught mice, we measured both the wild-caught mice and their laboratory-reared descendants for four traits previously reported to distinguish forest and prairie ecotypes (*7*–*9*): tail, hindfoot, and ear lengths as well as coat color (brightness, hue, and saturation across three body regions). We also measured body length and weight. We found that forest mice had longer tails; longer hind feet; and darker, redder coats compared with prairie mice (Fig. 1, B and C; fig. S1; and table S1). These phenotypic differences persisted in laboratory-born mice raised in common conditions (fig. S2 and table S1), which suggests a strong genetic component to these ecotype-defining traits.

A large inversion is associated with tail length and coat color

Using an unbiased forward-genetic approach, we identified genomic regions linked to ecotype differences in morphology. We intercrossed forest and prairie mice in the laboratory to generate 555 second-generation (F2) hybrids (forest female × prairie male, n = 203 F2s; prairie female × forest male, n = 352 F2s) and performed quantitative trait locus (QTL) mapping for each trait (*12*) (Fig. 2, fig. S3, and table S2). We identified five regions associated with tail length variation [total percent variance explained (PVE): 27%; individual PVE: 2.6 to 12.1%]. Only one region, on chromosome 15, was strongly and significantly associated with coat color variation (PVE, dorsal hue: 40.0%; PVE, flank hue: 45.6%). Each QTL exhibited incomplete dominance, and the forest allele was always associated with forest traits—longer tails or redder coats. The one significant QTL for coat color overlapped with the largest-effect locus associated with tail length (95% Bayesian credible intervals: dorsal hue = 0.4 to 40.5 Mb; flank hue = 0.4 to 39.4 Mb; tail length = 0.4 to 41.5 Mb). Thus, a single region on chromosome 15 was strongly associated with ecotype differences in both tail length and coat color.

The QTL peak on chromosome 15 exhibited a consistently strong association with both morphological traits across half the chromosome (Fig. 3A). This pattern reflects reduced recombination between forest and prairie alleles in the laboratory cross: Only 2 of 1110 F2 chromosomes were recombinant in this region (Fig. 3B). We also found consistently elevated F_{ST} (proportion of the total genetic variance explained by population structure) (Fig. 3C) and high linkage disequilibrium (Fig. 3D) across this genetic region in wild populations relative to the rest of the chromosome (whole-genome resequencing: n = 15 forest, n = 15 prairie). Together, these data are consistent with reduced recombination across half of chromosome 15 in both laboratory and wild populations.

This pattern of suppressed recombination could be produced by a large genomic rearrangement (or a set of rearrangements). To determine the nature of any structural variation on chromosome 15, we used PacBio long-read sequencing (n = 1 forest, n = 1 prairie) (*12*). We generated independent de novo assemblies for each individual and mapped the resulting contigs to the reference genome for *P. m. bairdii* (*12*). In the forest individual, one contig mapped near the center of the chromosome (from 41.19 to 40.94 Mb) and then split and mapped in reverse orientation to the beginning of the

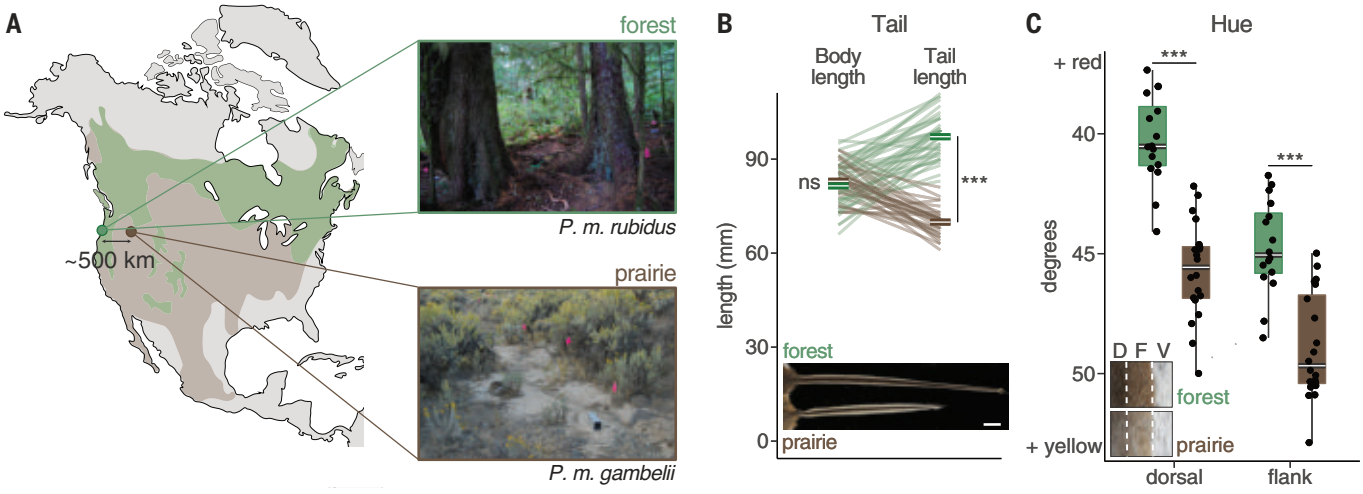


Fig. 1. Forest and prairie mice differ in tail length and pigmentation. (**A**) Map shows the approximate range of forest (green) and prairie (brown) deer mouse ecotypes in North America. Collection sites of wild-caught forest (*P. m. rubidus*, green) and prairie (*P. m. gambelii*, brown) ecotypes from western and eastern Oregon, USA, respectively, are shown. Photos illustrate representative habitat; pink flags indicate trap lines. (**B**) Body length (left; not including the tail) and tail length (right) for wild-caught adult mice (n = 38 forest and 32 prairie). Lines connect body and tail measurements for the same individual. Means are shown in bold. (Inset) Image of a representative tail from each ecotype. Scale bar, 1 cm. (**C**) Coat color (hue) values for the dorsal and flank regions of wild-caught adult mice (n = 16 forest and 20 prairie). Boxplots indicate the median (center white line) and the 25th and 75th percentiles (box extents); whiskers show largest or smallest value within 1.5 times the interquartile range. Black dots show individual data points. (Inset) Dorsal (D), flank (F), and ventral (V) regions from a representative forest and prairie mouse. ns = $P > 0.05$; ***$P < 0.001$ (Welch's t test, two-sided). Original photography in (B) and (C) is copyrighted by the President and Fellows of Harvard College (photo credit: Museum of Comparative Zoology, Harvard University).

chromosome (from 0 to 5 Mb). By contrast, in the prairie individual, a single contig mapped continuously to the reference genome in this region (37 to 41.3 Mb) (Fig. 3E). Because we found no other forest-specific rearrangements in this region (fig. S4), we determined that chromosome 15 harbors a simple 41-Mb inversion. Using putative centromere-associated sequences in *Peromyscus* (*12*), we determined that the inversion is paracentric, with the centromere located outside of the inversion (Fig. 3G).

Inversions may affect phenotypes directly through the effects of their breakpoints or indirectly by carrying causal mutations (*13*). Using the long-read sequencing data, we localized the inversion breakpoint to base pair resolution (Fig. 3F and fig. S5). The breakpoint falls within an intron of a long intergenic noncoding RNA (lincRNA), and an additional four annotated genes (two lincRNAs and two protein-coding genes) occur within 200 kb of the breakpoint. Although the breakpoint may disrupt their expression patterns, these genes have no known functions associated with either pigmentation or skeletal phenotypes (table S3). An additional 149 protein-coding genes are located within the inversion, of which 29 contain at least one fixed nonsynonymous mutation between the inversion and reference alleles. Ten of the genes within the inversion (four with nonsynonymous substitutions) are associated with pigmentation phenotypes when disrupted in laboratory mice, and 13 are associated with tail or long-bone length phenotypes in laboratory mice (three with nonsynonymous substitutions and four with associated pigment phenotypes as well; table S4). These 19 genes are thus strong candidates for contributing to tail length and coat color variation.

Inversion frequency and divergence in wild populations

To investigate whether the inversion and associated traits (longer tails and redder coats) may be favored in forested habitats, we collected deer mice across a sharp habitat transition between the focal forest and prairie sites and estimated habitat type and mean soil hue at each capture site (n = 136 mice from 22 sites, supplemented by 12 additional museum specimens from two sites; figs. S6 and S7). We found that much of the transition in both habitat type and soil hue occurs in a narrow region across the Cascade mountain range (Fig. 4, A and B), and the phenotypic clines estimated using either all adult wild-caught individuals or only those from the Cascades region both identified sharp transitions in coat color and tail length that colocalize with this environmental transition (Fig. 4, C and D). Specifically, mean hue changes by 3.2° (63% of the forest-prairie difference), and mean tail length changes by 13 mm (47% of the forest-prairie difference) across the 50-km Cascades region; tail length changes by an additional 4 mm within the next 100 km, coincident with continued changes in forestation (Fig. 4). Together, the strong correlation between phenotype and habitat is consistent with local adaptation.

The inversion changes substantially in frequency across the habitat transition, from 90% in the forest population to absent in the prairie population (Fig. 4E). This frequency difference of the inversion is extreme: It is greater than the allele frequency difference at the maximally differentiated single-nucleotide polymorphism (SNP) in 99.92% of blocks with similar levels of linkage disequilibrium (*12*) (Fig. 4F). Moreover, similar to the changes in phenotype, the transition in inversion frequency occurs over only a short distance: Inversion frequency decreases from 100 to 62.5% in the 50-km Cascades region and then drops further within the next 100 km (i.e., inversion frequency drops from 100 to 4% over less than one-third of the total transect distance; Fig. 4E). The sharp change in inversion frequency across the environmental transect, and its extreme forest-prairie allele frequency difference, suggest that the inversion may be favored in forested habitat.

The inversion also strongly contributes to genetic differentiation between the forest and prairie ecotypes by carrying many highly differentiated SNPs. For example, F_{ST} between the forest and prairie ecotypes in the inversion region is high compared with the genome-wide average (inversion region: mean F_{ST} = 0.376; genome-wide, excluding inversion region: mean F_{ST} = 0.071; fig. S8). The strong genetic divergence between the inversion and reference haplotypes is reflected in maximum likelihood–based trees built from the region of chromosome 15 that contains the inversion (affected region:

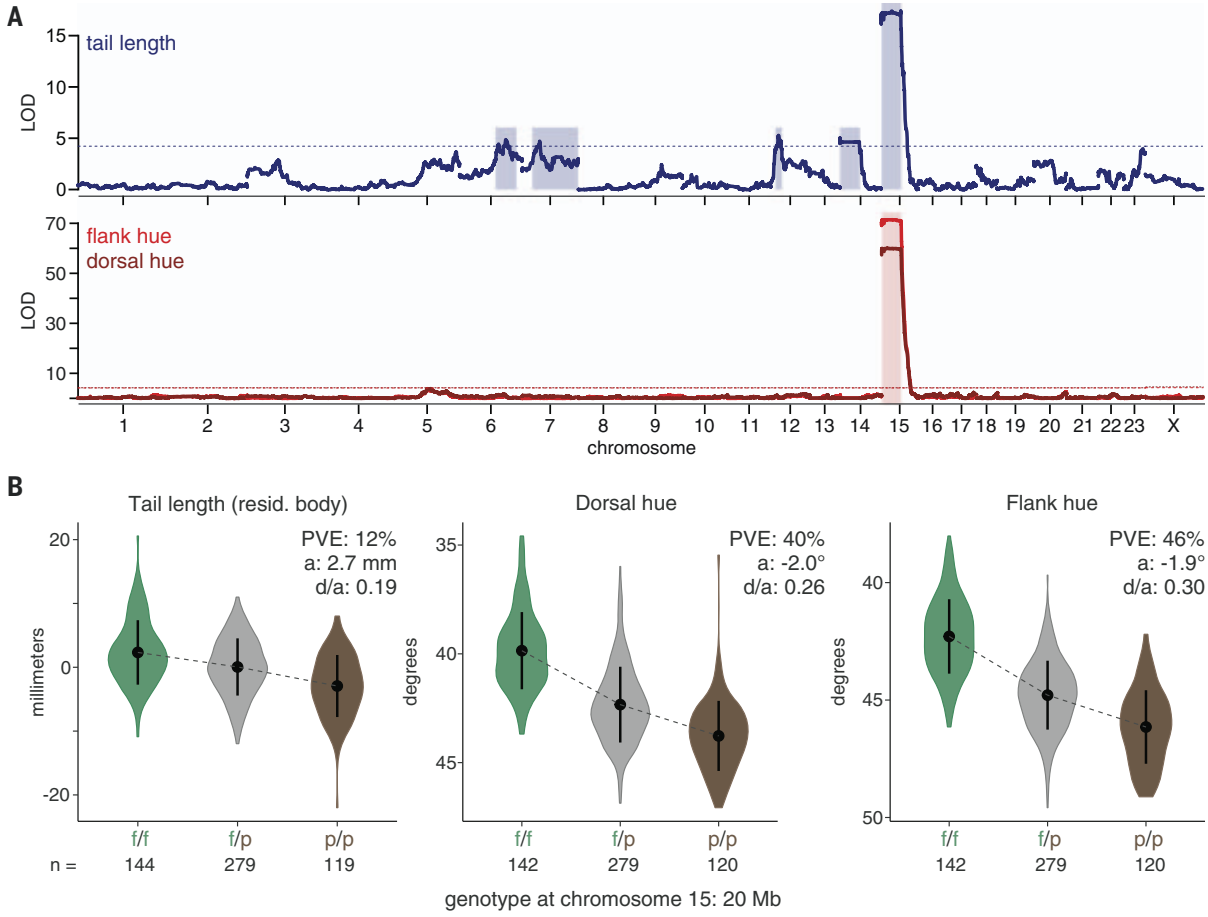


Fig. 2. A region on chromosome 15 is strongly associated with both tail length and coat color. (**A**) Statistical association [log of the odds (LOD) score] of ancestry with tail length (top; blue) and dorsal and flank hue (bottom; dorsal, dark red; flank, light red) in laboratory-reared F2 hybrids (tail, n = 542; hue, n = 541). Physical distance (in base pairs) is shown on the x axis; axis labels indicate the center of each chromosome. Dotted lines indicate the genome-wide significance threshold (α = 0.05) based on permutation tests, and shaded rectangles indicate the 95% Bayesian credible intervals for all chromosomes with significant QTL peaks. For tail length analysis, body length was included as an additive covariate. (**B**) Tail length (left; shown after taking the residual against body length in the hybrids), dorsal hue (center), and flank hue (right) of F2 hybrids, binned by genotype at 20 Mb on chromosome 15 (f/f, homozygous forest; f/p, heterozygous; p/p, homozygous prairie) (sample sizes are given below the x axes). Points and error bars show means ± standard deviations. PVE, percent of the variance explained by genotype; a, additive effect of one forest allele; d/a, absolute value of the dominance ratio.

0 to 40.9 Mb) and the rest of the chromosome (unaffected region: 40.9 to 79 Mb). In the unaffected region, forest and prairie mice cluster by ecotype, with limited divergence between the groups (Fig. 4G). By contrast, in the affected region, mice cluster into two highly distinct groups on the basis of genotypes at the inversion (Fig. 4H). This pattern suggests that the inversion harbors a high density of sites that are divergent between ecotypes.

Evolutionary history of the inversion

To explore the evolutionary history of the inversion, we first estimated a best-fitting demographic model for the forest and prairie populations using neutral sites across the genome to avoid the confounding effects of background selection (*12*, *14*). The data were best fit by a model with a long history of high migration: initial migration rates of 8.3×10^{-7} [prairie-to-forest, 95% confidence interval (CI) = 3.7×10^{-9} to 1.8×10^{-6}] and 3.6×10^{-6} (forest-to-prairie, 95% CI = 1.1×10^{-8} to 4.5×10^{-6}) after a forest-prairie population split 2.2 million generations ago (95% CI = 1.1 to 5.5 million generations) (Fig. 5A and fig. S9). Because the estimated effective population sizes (N_e) are large (prairie $N_e = 1.9 \times 10^6$ to 4.3×10^6; forest $N_e = 1.8 \times 10^5$ to 1.2×10^6), the effective number of migrants per generation (N_em) is consistently high over time: N_em = 3.5 (prairie-to-forest) and N_em = 0.6 (forest-to-prairie), with a recent shift to $N_em > 10$ in both directions ~30,000 generations ago (Fig. 5A), consistent with high levels of gene flow (*15*). High migration levels between forest and prairie ecotypes are further supported by genomic data from the Cascades region: We found that the Cascades mice have mixed forest and prairie ancestry genome-wide (fig. S10).

These high migration estimates coupled with the large, habitat-associated differences in inversion frequency may indicate a history of natural selection. To test this hypothesis, we simulated the spread of the inversion under our demographic model using SLiM (*12*). We found that divergent selection was the most likely scenario to explain both the high frequency of the inversion in the forest and its low frequency in the prairie (fig. S11). Using approximate Bayesian computation, we estimated selection coefficients (s) for the inversion of 3.3×10^{-4} (95% CI = 9.2×10^{-5} to 1.6×10^{-3}) in the forest population and -4.1×10^{-3} (95% CI = -9.3×10^{-3} to -7.1×10^{-4}) in the prairie population (Fig. 5B). These values suggest that the observed distribution of the inversion in

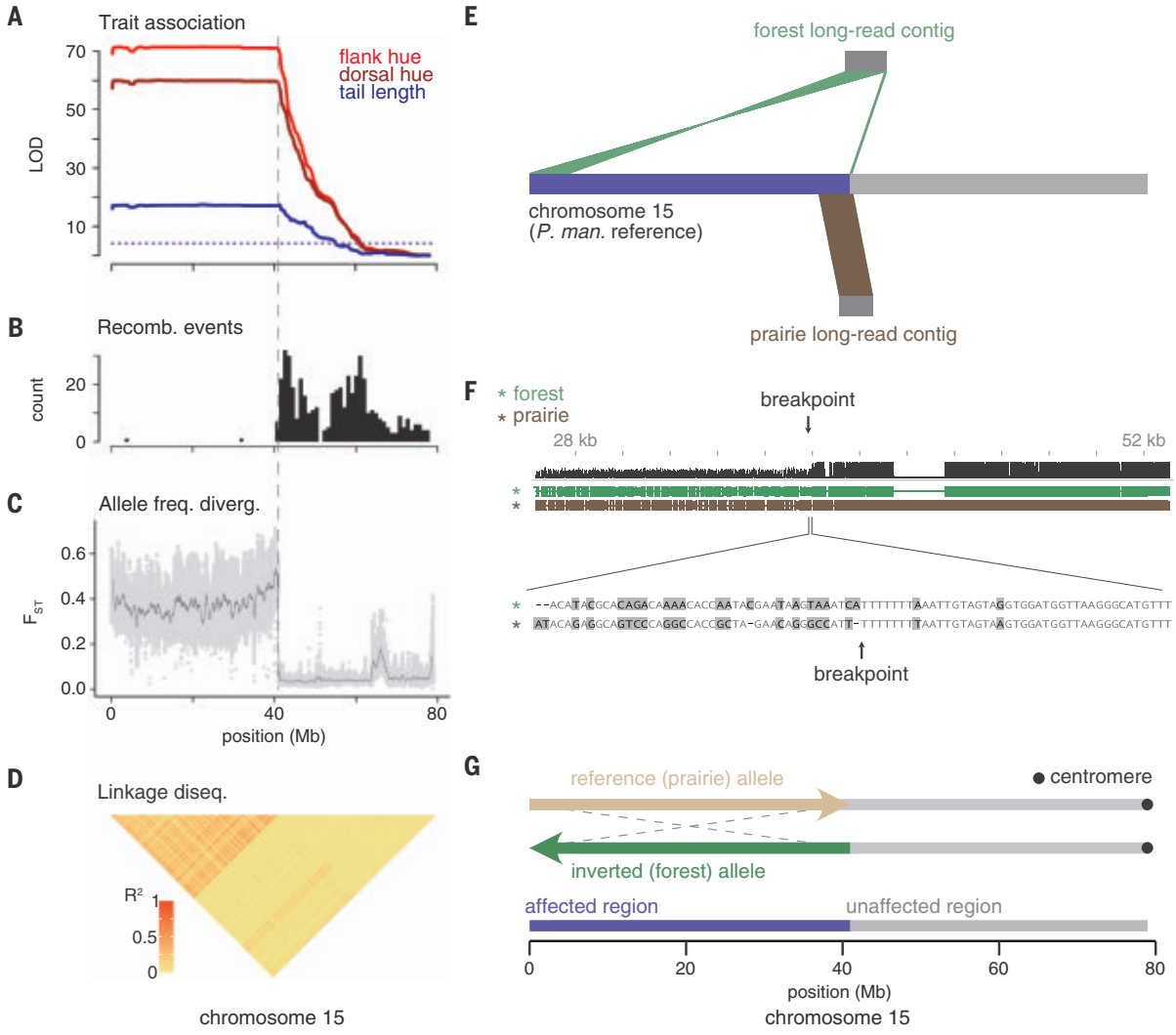


Fig. 3. Chromosomal region associated with tail length and coat color is a large inversion. Across chromosome 15, data are from F2 hybrids [(A) and (B)] and wild-caught mice [(C) and (D), (n = 15 forest and 15 prairie)]. (**A**) LOD score for tail length (blue), dorsal hue (dark red), and flank hue (light red). (**B**) Number of recombination breakpoint events, binned in 1-Mb windows. (**C**) F_{ST} between forest and prairie mice estimated in 10-kb windows with a step size of 1 kb (light gray dots). Dark gray line shows data smoothed with a moving average over 500 windows. (**D**) Linkage disequilibrium across forest and prairie mice. Heatmap shows R^2 (squared correlation) computed between genotypes at thinned SNPs (*12*). (**E**) Contigs assembled from long-read sequencing for one forest (top) and one prairie (bottom) mouse. Only contigs that span the inversion breakpoint are shown. The region of chromosome 15 affected by the inversion is highlighted (purple). (**F**) (Top) Alignment between regions of the forest and prairie contigs surrounding the breakpoint (black, alignment quality; green, forest contig; brown, prairie contig). Large prairie insertion near the breakpoint is a transposon. (Bottom) Base pair–level alignment around the breakpoint (gray, mismatch). (**G**) Model of the inverted (green) and reference (tan) alleles. The inversion spans 0 to 40.9 Mb (affected region, purple) and excludes 40.9 to 79 Mb (unaffected region, gray), with predicted centromere location shown in black.

the wild is best explained by both positive selection in the forest and negative selection in the prairie, a conclusion robust to the uncertainty in the model parameter estimates (fig. S12) and to variation in the timing of the introduction of the inversion after the forest-prairie split (fig. S13). We also used simulations to assess the minimum age of the inversion required to achieve its divergence from the reference allele (*12*): We estimated the inversion to be at least 247,000 generations old (95% CI = 149,000 to 384,000 generations or 50,000 to 128,000 years, assuming three generations per year), which suggests that the inversion predates the modern habitat distribution (*16*) (Fig. 5C). Together, these results suggest that the inversion was most likely established in the forest population under strong divergent selection over the last ~250,000 generations.

Our estimates of forest-prairie migration rates and selection on the inversion allowed us to explore possible fitness effects from the inversion's suppression of recombination. Although it is formally possible that the inversion carries only a single mutation that alone confers a strong enough benefit ($s \geq 3 \times 10^{-4}$) to explain its current distribution, an alternative hypothesis is that the inversion carries two or more beneficial mutations (e.g., one mutation that contributes to tail length and a second to color variation), each with smaller selection coefficients. In this scenario, theory predicts that the inversion could confer a fitness advantage in the forest beyond the individual mutations it carries by reducing the migration

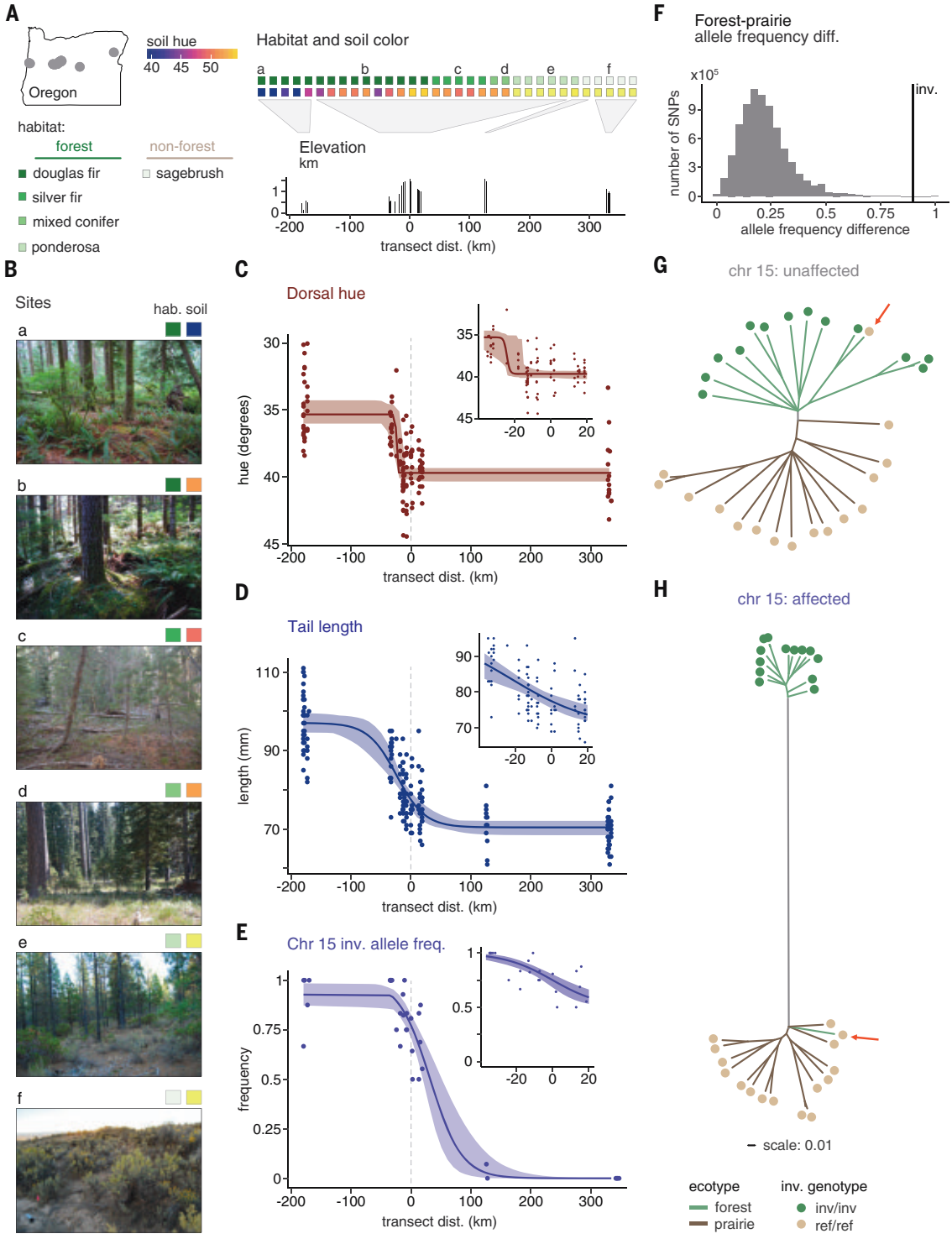


Fig. 4. Associations between genotype, phenotype, and environment in wild mice. (**A**) Elevation and habitat characteristics (top row indicates majority habitat category, and bottom row indicates mean soil hue) at sites across an environmental transect. Letters indicate sites shown in (B). Soil hue and habitat category were estimated within 1 km of each site. (Map) Sampled sites across Oregon. Transect distance refers to the east-west distance from the highest-elevation site, and dotted lines in (C), (D), and (E) indicate distance = 0. (**B**) Photos of capture sites from each habitat type, with habitat and soil classification as in (A). (**C** to **E**) Best-fit clines for dorsal hue (C) (n = 143), tail length (D) (n = 180), and inversion genotype (E) (n = 178) fit to the full dataset, with 95% CIs. Insets show best-fit clines using only data from the central Cascades (hue, n = 90; tail, n = 97; genotype, n = 136). (**F**) Allele frequency differences for the maximally differentiated SNP between forest and prairie mice in 200-bp windows across the genome (*12*). The inversion forest-prairie allele frequency difference (90%) is shown in black. (**G** and **H**) Maximum likelihood trees for unaffected (G) (40.9 to 79 Mb) and affected (H) (0 to 40.9 Mb) regions of chromosome 15, shown on the same scale. Branch colors indicate ecotype (green, forest; brown, prairie), and dots indicate inversion genotype (tan, homozygous reference, n = 15; green, homozygous inversion, n = 14; heterozygous mouse excluded, n = 1). Red arrows highlight the forest mouse homozygous for the reference allele.

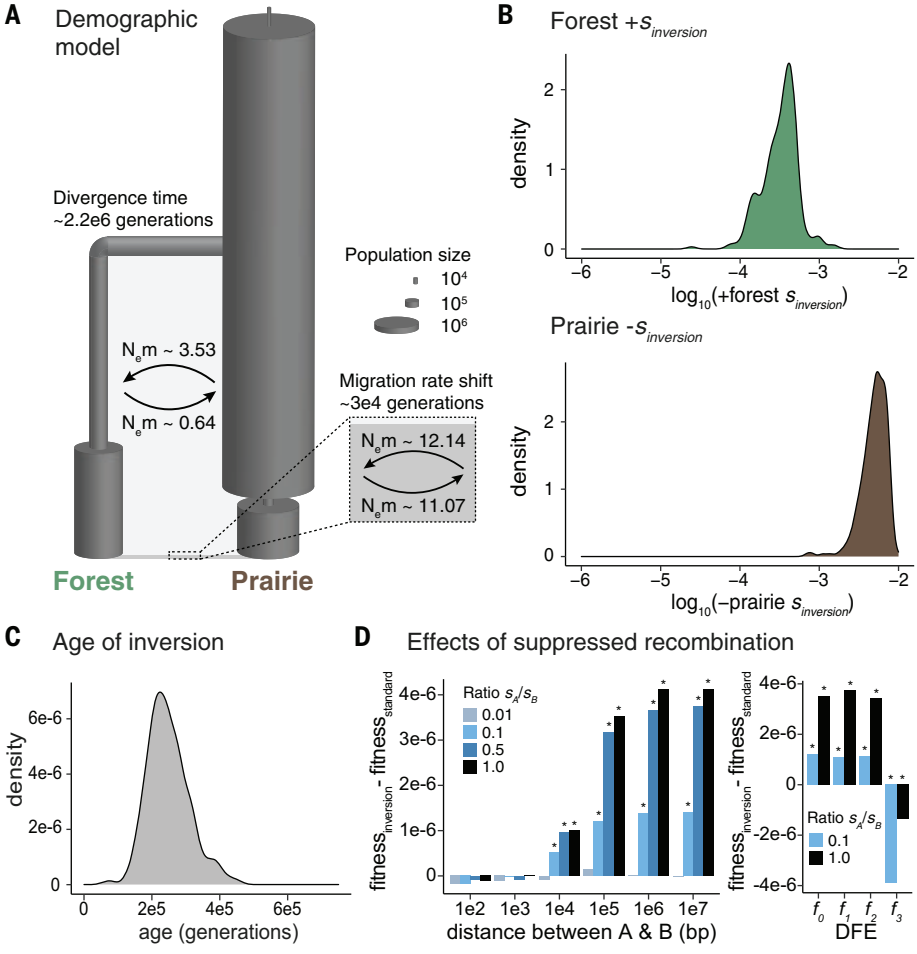


Fig. 5. Evolutionary history of the inversion. (**A**) Best-fit demographic model. N_e, effective population size; m, migration rate. (**B**) Posterior probability distributions for the selection coefficient associated with the inversion in the forest (top, green) and prairie (bottom, brown) populations, when the inversion is introduced 150,000 generations ago (for additional introduction times, see fig. S13). The estimated selection coefficient is positive in forest and negative in prairie. (**C**) Posterior probability distribution for the age of the inversion. (**D**) Estimated fitness effects of suppressed recombination within the inversion. Two beneficial loci (A and B) were introduced into the forest population on the inversion or on a standard haplotype, varying the ratio of the selection coefficients for A (s_A) and B (s_B), with $s_A + s_B$ kept constant at 3×10^{-4}. bp, base pairs. Bar height shows the difference in final mean fitness of the forest population between the inversion and standard haplotype scenarios. Asterisks indicate a significant difference in mean fitness ($P < 0.05$) computed with permutation tests. (Left) Two beneficial loci at varying distances apart, without deleterious mutations. (Right) Two beneficial loci separated by 100 kb, with deleterious mutations introduced according to distributions of fitness effects (DFE): f_0: 100% of mutations neutral ($2Ns = 0$, where N indicates population size and s indicates selection coefficient); f_1: 50% of mutations neutral ($2Ns = 0$), 50% weakly deleterious ($-10 < 2Ns < -1$); f_2: 33% of mutations neutral ($2Ns = 0$), 33% weakly deleterious ($-10 < 2Ns < -1$), 33% moderately deleterious ($-100 < 2Ns < -10$); f_4: 25% of mutations neutral ($2Ns = 0$), 25% weakly deleterious ($-10 < 2Ns < -1$), 25% moderately deleterious ($-100 < 2Ns < -10$), 25% strongly deleterious ($-1000 < 2Ns < -100$).

load suffered by each mutation (*5*, *17*, *18*). To investigate this possibility, we used our estimates of migration, selection, and recombination to simulate the spread of two beneficial mutations in the forest population either within an inversion or on a freely recombining (standard) haplotype, varying the distance between the mutations (*12*). We found that if the two mutations are at least 10 kb apart (which is likely, given the inversion size of 41 Mb) and the selection coefficient for the weaker locus is at least 10% of that of the stronger locus [which is possible, given independent evidence for selection acting on coat color and tail length—e.g., (*19*, *20*)], the beneficial mutations are more likely to establish and be maintained at higher frequencies in the forest when carried by the inversion than on the standard haplotype (Fig. 5D and figs. S14 and S15). We also explored possible costs associated with the inversion suppressing recombination (i.e., mutational load accumulation) (*21*, *22*) by introducing deleterious mutations according to four fitness-effect distributions [as described in (*14*)] into the two–beneficial locus simulations. With weakly or moderately deleterious mutations, the inversion maintained its selective advantage over the standard haplotype in the forest (Fig. 5D and fig. S16). Only when strongly deleterious mutations were introduced did the inversion accumulate a substantial mutational load, which results in the inversion being disadvantageous relative to the standard haplotype in the forest (Fig. 5D and fig. S16). Thus, our results suggest that, under a wide range of conditions, if this inversion carries two or more beneficial mutations, its suppression of recombination likely confers an additional selective advantage in the forest population by linking adaptive alleles in the face of high migration rates.

Discussion

In 1909, Wilfred Osgood described several morphological differences—including tail length and coat color—that distinguish forest and prairie ecotypes of *P. maniculatus* (*7*). Long tails are thought to be beneficial for arboreality (*8*, *9*, *23*): Long tails have repeatedly evolved in association with forest habitat in deer mice (*20*) and across mammals (*24*), and forest mice are better climbers (*23*), with tail length differences between the ecotypes likely sufficient to affect climbing performance (*25*). Coat color is subject to pressure from visually hunting predators (*19*), and many mammals, including deer mice, evolve coats to match local soil color (*9*, *26*). By sampling along an environmental transect, we found evidence that each of these traits is closely associated with habitat (forestation for tail length and soil hue for coat color), which further suggests that these traits are involved in local adaptation.

High migration rates between the forest and prairie ecotypes, as we estimated in this work, makes the strong ecotypic divergence in multiple traits puzzling. By characterizing the genetic architecture of tail length and coat color variation, we help resolve how differences in these traits are maintained between ecotypes: Namely, we discover a previously unknown inversion, involving half a chromosome, that has a large effect on both ecotype-defining

traits and in the expected direction (i.e., it is associated with long tails and reddish fur in forest mice). Because recombination between the inversion and the noninverted prairie haplotype is suppressed in heterozygotes, the inversion ensures that longer tail length and redder coat color alleles are coinherited in the forest, despite high levels of gene flow (except in the unlikely scenario that only a single pleiotropic mutation within the inversion affects both traits). The role of this inversion in phenotypically differentiating these ecotypes is consistent with theoretical predictions and empirical examples of concentrated genetic architectures arising under local adaptation with gene flow (*6*, *27*, *28*).

Our modeling implicates divergent selection in maintaining the inversion at high frequency in the forest ecotype and absent from the prairie ecotype. The inversion's selective effects are likely driven by its strong association with tail length and coat color (explaining 12 and 40% of the trait variances, respectively), although it is possible other traits are involved. Although inversions can have phenotypic effects because of their breakpoints disrupting genes or gene expression (*13*), the inversion's breakpoint does not occur in or near candidate genes for tail length and coat color variation. Alternatively, inversions may influence phenotypes through the mutations they carry: The inversion is highly differentiated from the reference haplotype, thus harboring many mutations that may influence tail length and/or coat color. We expect that more than one mutation contributes to the inversion's selective benefit in the forest, given the size of the inversion (41 Mb), its large selection coefficient in the forest ($s \approx 3 \times 10^{-4}$, or $Ns \approx 120$), and its association with two largely developmentally distinct traits. If this is the case, the inversion's suppression of recombination likely provides an additional benefit (beyond the individual effects of its mutations) in the forest population, as long as strongly deleterious mutations are uncommon. This finding—that recombination suppression is likely beneficial in this system—provides empirical support for the local adaptation hypothesis, which posits that inversions are beneficial in the face of gene flow because they increase linkage disequilibrium between adaptive alleles (*5*, *17*, *18*).

One hundred years after Alfred Sturtevant first provided evidence of chromosomal inversions in laboratory stocks of *Drosophila* (*29*) and, separately, forest-prairie ecotypes were first described in wild populations of *Peromyscus* (*7*), we found that a large chromosomal inversion is key to ecotype divergence in this classic system. Inversions have been identified in association with divergent ecotypes in diverse species, including plants (*30*–*33*), invertebrates (*34*–*45*), fish (*46*, *47*), and birds (*48*–*52*). In mammals, however, evidence for ecotype-defining inversions is limited [(*53*), but see (*54*)]. Our results thus underscore the important and perhaps widespread role of inversions in local adaptation, including in mammals, and highlight how selection acting on inversion polymorphisms may maintain intraspecific divergence in multiple traits in the wild.

REFERENCES AND NOTES

1. G. Turesson, *Hereditas* **3**, 100–113 (1922).
2. D. B. Lowry, *Biol. J. Linn. Soc.* **106**, 241–257 (2012).
3. A. Tigano, V. L. Friesen, *Mol. Ecol.* **25**, 2144–2164 (2016).
4. J. B. S. Haldane, *Math. Proc. Camb. Philos. Soc.* **26**, 220–230 (1930).
5. R. Bürger, A. Akerman, *Theor. Popul. Biol.* **80**, 272–288 (2011).
6. S. Yeaman, M. C. Whitlock, *Evolution* **65**, 1897–1911 (2011).
7. W. H. Osgood, *N. Am. Fauna* **28**, 1–285 (1909).
8. W. F. Blair, *Evolution* **4**, 253–275 (1950).
9. L. R. Dice, *Am. Nat.* **74**, 212–221 (1940).
10. S. W. Calhoun, I. F. Greenbaum, K. P. Fuxa, *J. Mammal.* **69**, 34–45 (1988).
11. D. S. Yang, G. Kenagy, *Ecol. Evol.* **1**, 26–36 (2011).
12. Materials and methods are available as supplementary materials.
13. R. Villoutreix *et al*., *Mol. Ecol.* **30**, 2738–2755 (2021).
14. P. Johri *et al*., *Mol. Biol. Evol.* **38**, 2986–3003 (2021).
15. S. Wright, *Genetics* **16**, 97–159 (1931).
16. G. Hewitt, *Nature* **405**, 907–913 (2000).
17. M. Kirkpatrick, N. Barton, *Genetics* **173**, 419–434 (2006).
18. B. Charlesworth, N. H. Barton, *Genetics* **208**, 377–382 (2018).
19. S. N. Vignieri, J. G. Larson, H. E. Hoekstra, *Evolution* **64**, 2153–2158 (2010).
20. E. P. Kingsley, K. M. Kozak, S. P. Pfeifer, D. S. Yang, H. E. Hoekstra, *Evolution* **71**, 261–273 (2017).
21. E. L. Berdan, A. Blanckaert, R. K. Butlin, C. Bank, *PLOS Genet.* **17**, e1009411 (2021).
22. B. Charlesworth, J. D. Jensen, *Annu. Rev. Ecol. Evol. Syst.* **52**, 177–197 (2021).
23. B. E. Horner, *Contrib. Lab. Vertebr. Biol.* **61**, 1–85 (1954).
24. S. T. Mincer, G. A. Russo, *Proc. R. Soc. B.* **287**, 20192885 (2020).
25. E. R. Hager, H. E. Hoekstra, *Integr. Comp. Biol.* **61**, 385–397 (2021).
26. R. D. H. Barrett *et al*., *Science* **363**, 499–504 (2019).
27. R. Faria, K. Johannesson, R. K. Butlin, A. M. Westram, *Trends Ecol. Evol.* **34**, 239–248 (2019).
28. M. J. Thompson, C. D. Jiggins, *Heredity* **113**, 1–8 (2014).
29. A. H. Sturtevant, *Proc. Natl. Acad. Sci. U.S.A.* **7**, 235–237 (1921).
30. Z. Fang *et al*., *Genetics* **191**, 883–894 (2012).
31. K. Huang, R. L. Andrew, G. L. Owens, K. L. Ostevik, L. H. Rieseberg, *Mol. Ecol.* **29**, 2535–2549 (2020).
32. D. B. Lowry, J. H. Willis, *PLOS Biol.* **8**, e1000500 (2010).
33. C.-R. Lee *et al*., *Nat. Ecol. Evol.* **1**, 0119 (2017).
34. M. Joron *et al*., *Nature* **477**, 203–206 (2011).
35. K. Kunte *et al*., *Nature* **507**, 229–232 (2014).
36. Z. Yan *et al*., *Nat. Ecol. Evol.* **4**, 240–249 (2020).
37. C. Mérot *et al*., *Mol. Biol. Evol.* **38**, 3953–3971 (2021).
38. C. Cheng *et al*., *Genetics* **190**, 1417–1432 (2012).
39. J. L. Feder, J. B. Roethele, K. Filchak, J. Niedbalski, J. Romero-Severson, *Genetics* **163**, 939–953 (2003).
40. M. Kapun, T. Flatt, *Mol. Ecol.* **28**, 1263–1282 (2019).
41. E. L. Koch *et al*., *Evol. Lett.* **5**, 196–213 (2021).
42. D. Lindtke *et al*., *Mol. Ecol.* **26**, 6189–6205 (2017).
43. T. Dobzhansky, *Evolution* **1**, 1–16 (1947).
44. A. Brelsford *et al*., *Curr. Biol.* **30**, 304–311.e4 (2020).
45. D. Ayala, R. F. Guerrero, M. Kirkpatrick, *Evolution* **67**, 946–958 (2013).
46. F. C. Jones *et al*., *Nature* **484**, 55–61 (2012).
47. M. Matschiner *et al*., *Nat. Ecol. Evol.* **6**, 469–481 (2022).
48. C. Küpper *et al*., *Nat. Genet.* **48**, 79–83 (2016).
49. H. B. Thorneycroft, *Science* **154**, 1571–1572 (1966).
50. S. Lamichhaney *et al*., *Nat. Genet.* **48**, 84–88 (2016).
51. E. R. Funk *et al*., *Nat. Commun.* **12**, 6833 (2021).
52. I. Sanchez-Donoso *et al*., *Curr. Biol.* **32**, 462–469.e6 (2022).
53. G. Dobigny, J. Britton-Davidian, T. J. Robinson, *Biol. Rev.* **92**, 1–21 (2017).
54. H. Stefansson *et al*., *Nat. Genet.* **37**, 129–137 (2005).
55. E. R. Hager, O. S. Harringmeyer, oharring/chr15_inversion: v1.0.0, version 1.0.0, Zenodo (2022); https://doi.org/10.5281/zenodo.6354918.

ACKNOWLEDGMENTS

We thank C. Lewarch for assistance in the field; S. Niemi, M. Streisfeld, S. Stankowski, G. Binford, S. Bishop, K. Saunders, S. Finch, and T. Schaller for help with field logistics; K. Pritchett-Corning and S. Griggs-Collette for help establishing breeding colonies; Harvard's Office of Animal Resources for animal care; and M. Omura, J. Chupasko, M. Mullon, and J. Mewherter for help preparing and accessioning museum specimens. G. J. Kenagy, D. S. Yang, and E. Kingsley provided advice at the start of the project; P. Audano provided advice on long-read sequencing and analysis; and N. Edelman, A. Kautt, E. Kingsley, and three anonymous reviewers provided helpful feedback on the manuscript. The University of Washington Burke Museum provided specimens used in this study. **Funding:** This work was partially funded by Putnam Expedition grants from the Museum of Comparative Zoology (MCZ) to E.R.H., a MCZ grant-in-aid of student research to S.T., and the MCZ Chapman Fellowship for the Study of Vertebrate Locomotion to E.R.H. and J.T.G. as well as funding from the American Society of Mammalogists grants-in-aid of research to E.R.H. and O.S.H., the Harvard College Research Program to S.T., and a Society for the Study of Evolution R. C. Lewontin Early Award to O.S.H. E.R.H. was supported by an NIH training grant to Harvard's Molecules, Cells, and Organisms graduate program (NIH NIGMS T32GM007598) and by the Theodore H. Ashford Fellowship. O.S.H. was supported by a National Science Foundation (NSF) Graduate Research Fellowship, a Harvard Quantitative Biology Student Fellowship (DMS 1764269), and the Molecular Biophysics training grant (NIH NIGMS T32GM008313). J.D.J. was funded by the National Institutes of Health (1R35GM139383-01). H.E.H. is an investigator of the Howard Hughes Medical Institute. **Author contributions:** E.R.H. and H.E.H. initially conceived of the project. E.R.H., J.T.G., and K.M.T. planned and conducted the field collections. E.R.H. conducted the QTL mapping experiment and analyzed the forest-prairie phenotypes; E.R.H. and K.M.T. generated genetic data for the cross; and E.R.H., J.T.G., S.T., S.M., B.N., and K.M.T. generated wild and laboratory phenotype data. O.S.H. analyzed genetic data for the F2 cross, E.R.H. and O.S.H. performed QTL mapping analyses, T.B.W. analyzed long-read sequence data, O.S.H. and T.B.W. analyzed wild-caught forest-prairie genetic data, O.S.H. analyzed transect genetic data, and E.R.H. and O.S.H. performed cline analyses. T.B.W. performed demographic simulation, model-fitting, and inference, and O.S.H. performed selection simulations, model-fitting, and inference, both with input from J.D.J. E.R.H., O.S.H., T.B.W., J.D.J., and H.E.H. wrote the manuscript, with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Associated data are available as supplementary files data S1 to S8 and on the NCBI Sequence Read Archive (nos. PRJNA687993, PRJNA688305, and PRJNA816517). Associated scripts are available on Zenodo (*55*). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg0718
Materials and Methods
Figs. S1 to S21
Tables S1 to S6
References (*56*–*95*)
MDAR Reproducibility Checklist
Data S1 to S8

Submitted 14 December 2020; resubmitted 26 November 2021
Accepted 24 June 2022
10.1126/science.abg0718

REPORTS

CATALYSIS

Physical mixing of a catalyst and a hydrophobic polymer promotes CO hydrogenation through dehydration

Wei Fang[1]†, Chengtao Wang[1,2]†, Zhiqiang Liu[3]†, Liang Wang[1]*, Lu Liu[1], Hangjie Li[2], Shaodan Xu[4], Anmin Zheng[3]*, Xuedi Qin[2], Lujie Liu[1], Feng-Shou Xiao[1,5]*

In many reactions restricted by water, selective removal of water from the reaction system is critical and usually requires a membrane reactor. We found that a simple physical mixture of hydrophobic poly(divinylbenzene) with cobalt-manganese carbide could modulate a local environment of catalysts for rapidly shipping water product in syngas conversion. We were able to shift the water-sorption equilibrium on the catalyst surface, leading to a greater proportion of free surface that in turn raised the rate of syngas conversion by nearly a factor of 2. The carbon monoxide conversion reached 63.5%, and 71.4% of the hydrocarbon products were light olefins at 250°C, outperforming poly(divinylbenzene)-free catalyst under equivalent reaction conditions. The physically mixed CoMn carbide/poly(divinylbenzene) catalyst was durable in the continuous test for 120 hours.

Selective and rapid removal of water product from a reaction system has been a highly desirable pathway toward boosting catalytic performance in reactions that are restricted by water thermodynamically and/or kinetically (*1*, *2*). Membrane reactors designed to include water-conduction nanochannels could shift the reaction equilibrium (*3*), but preparation of defect-free membranes at a large scale is challenging. Chemical hydrophobilization of the catalyst surface could substantially contribute to reactions by accelerating water diffusion (*4*–*7*), but in many cases the chemical interactions might change the structure of the catalyst surface or even block the active sites by hindering access of reactant molecules.

Enabling rapid water diffusion from an unchanged catalyst surface is an attractive alternative. By promoting rapid desorption of water molecules once they are formed on the catalyst surface (*2*), the sorption equilibrium of water is shifted, as described by $*H_2O \rightleftharpoons * + H_2O$ (* denotes the catalyst surface sites). We physically mixed the hydrophobic promoter with the catalyst, unlike previous chemical modifications; in the syngas conversion to olefins with a cobalt-manganese carbide (CoMnC) catalyst, we achieved an increase in light olefin productivity by a factor of up to 3.4 by mixing the catalyst with the promoter. Mechanistic studies revealed that the water molecules rapidly desorbed from the catalyst surface after they formed from CO hydrogenation, which avoided the competitive adsorption with CO reactant, a crucial step in the reaction process.

The CoMnC catalyst was prepared via coprecipitation and carbonization procedures (fig. S1) (*8*, *9*). In the syngas conversion under the given reaction conditions (H_2/CO of 2, 1800 ml g_{CoMnC}^{-1} $hour^{-1}$, 0.1 MPa, 250°C), the CoMnC catalyst showed a CO conversion of 32.2%, with selectivity for light olefins ($C_2^=$ to $C_4^=$) of 60.8% (fig. S2 and table S1; the CO_2 product was excluded in calculating the selectivity). In our initial attempt, we physically mixed the CoMnC catalyst with a nonporous poly(divinylbenzene) (PDVB) (water-droplet contact angle 145°, irregular morphology, surface area <5 m^2/g; table S2). This hydrophobic polymer has a chemically inert surface (fig. S3) and good thermal stability (*10*). The mixed catalyst, denoted CoMnC/PDVB, had substantially improved CO conversion (63.5%) and selectivity to light olefins (71.4%) (Fig. 1) relative to the CoMnC catalyst under equivalent reaction conditions. With regard to the C_{5+} by-products over CoMnC/PDVB, the proportions of pentene and hexene were 50.3% and 26.0%, respectively (table S3); these are desired products for the production of high-performance polymers and valuable chemicals (*11*, *12*).

In particular, the molar ratios of olefin to paraffin (*o*/*p*) in these by-products were very high. For example, the *o*/*p* ratio of C_6 molecules is 14.7, which is favorable for separation and purification of desired products (*13*). A higher gas hourly space velocity (GHSV) of 3600 ml g_{CoMnC}^{-1} $hour^{-1}$ decreased the CO conversion to 15.4% on the CoMnC catalyst, but the CoMnC/PDVB still showed a remarkably high CO conversion of 50.7% (table S1). Under a GHSV of 7200 ml g_{CoMnC}^{-1} $hour^{-1}$, the CoMnC catalyst exhibited poor CO conversion of 5.5%, whereas the CoMnC/PDVB still showed CO conversion of 19.4%. In this case, the space-time productivity of light olefins on the CoMnC/PDVB reached as high as 7.1 mmol g_{CoMnC}^{-1} $hour^{-1}$, which exceeded the rate for CoMnC (2.1 mmol g_{CoMnC}^{-1} $hour^{-1}$, carbon basis) by a factor of 3.4 (fig. S4). In addition, a CO_2 selectivity of 46.0 to 48.5% was obtained in these cases, similar to that in general Fischer-Tropsch synthesis to olefins (*8*, *14*) and OX-ZEO (oxide-zeolite) reaction processes for converting syngas to olefins (*15*, *16*). Adjusting the H_2/CO ratio to 3.6 and introducing a small amount of CO_2 in the syngas feed reduced the CO_2 selectivity in the products to 23.7%, giving a one-pass yield of light olefins at 28.0% (CO_2 included in calculating the olefin yield; table S4).

We compared the performance of CoMnC/PDVB to different catalysts tested previously in syngas conversion to light olefins. The data in fig. S5 show the comparison of the selectivity-conversion results reported for various catalysts in their stable period during the reaction (table S5). The general Fischer-Tropsch synthesis to olefin processes produced olefins having a wide carbon number distribution within the range C_1–C_{20} (*17*, *18*). The selectivity of 53.0% to light olefins, with a CO conversion of 80.0% over Fe-based catalysts, has been reported as one of the most efficient processes (*14*), requiring a high reaction temperature of 340°C. The OX-ZEO process (*15*), which combines the cascade reactions of CO hydrogenation over metal oxide and C-C coupling of methanol or ketene intermediates over zeolite, showed superior selectivity to light olefins but yielded low CO conversion (e.g., 17.0% over the $ZnCrO_x$/SAPO-34 catalyst) at even higher temperatures (400°C). Relative to these processes, the reaction over the CoMnC/PDVB catalyst proceeded at lower reaction temperatures with high selectivity to light olefins and efficiently suppressed methane formation. The CoMnC/PDVB catalyst also exhibited enhanced catalytic performance relative to the bare CoMn catalyst, which has been regarded as a superior catalyst for low-temperature syngas conversion to light olefins (*8*).

The catalytic performance of CoMnC/PDVB catalyst was influenced by the manner of mixing of the CoMnC and PDVB components (Fig. 2, A and B, table S6, and fig. S6). Compared with a

[1]Key Lab of Biomass Chemical Engineering of Ministry of Education, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310027, China. [2]Key Lab of Applied Chemistry of Zhejiang Province, Department of Chemistry, Zhejiang University, Hangzhou 310028, China. [3]National Center for Magnetic Resonance in Wuhan, State Key Laboratory of Magnetic Resonance and Atomic and Molecular Physics and Mathematics, Wuhan Institute of Physics and Mathematics, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan 430071, China. [4]College of Materials and Environmental Engineering, Hangzhou Dianzi University, Hangzhou 310018, China. [5]Beijing Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, China.
***Corresponding author. Email: liangwang@zju.edu.cn (L.W.); zhenganm@wipm.ac.cn (A.Z.); fsxiao@zju.edu.cn (F.-S.X.)**
†These authors contributed equally to this work.

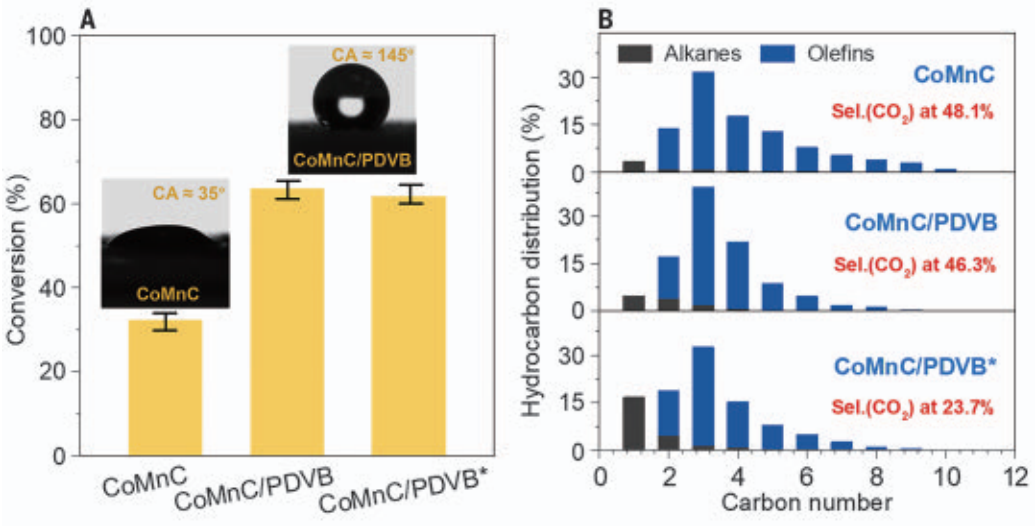


Fig. 1. Catalytic data in light olefin production from syngas. (**A** and **B**) CO conversion (A) and hydrocarbon distribution (B) over the CoMnC and CoMnC/PDVB catalysts. Reaction conditions: 1.0 g of CoMnC catalyst or mixture containing 1.0 g of CoMnC and 1.0 g of PDVB, H_2/CO/Ar at 64/32/4, 1800 ml g_{CoMnC}^{-1} $hour^{-1}$, 0.1 MPa, 250°C. The error bounds were estimated by repeating the experiment more than six times. Inset in (A): Water-droplet contact angles (CAs) of CoMnC and CoMnC/PDVB. *Reaction using a feed gas of H_2/CO/Ar/CO_2 at 68/19/3/10.

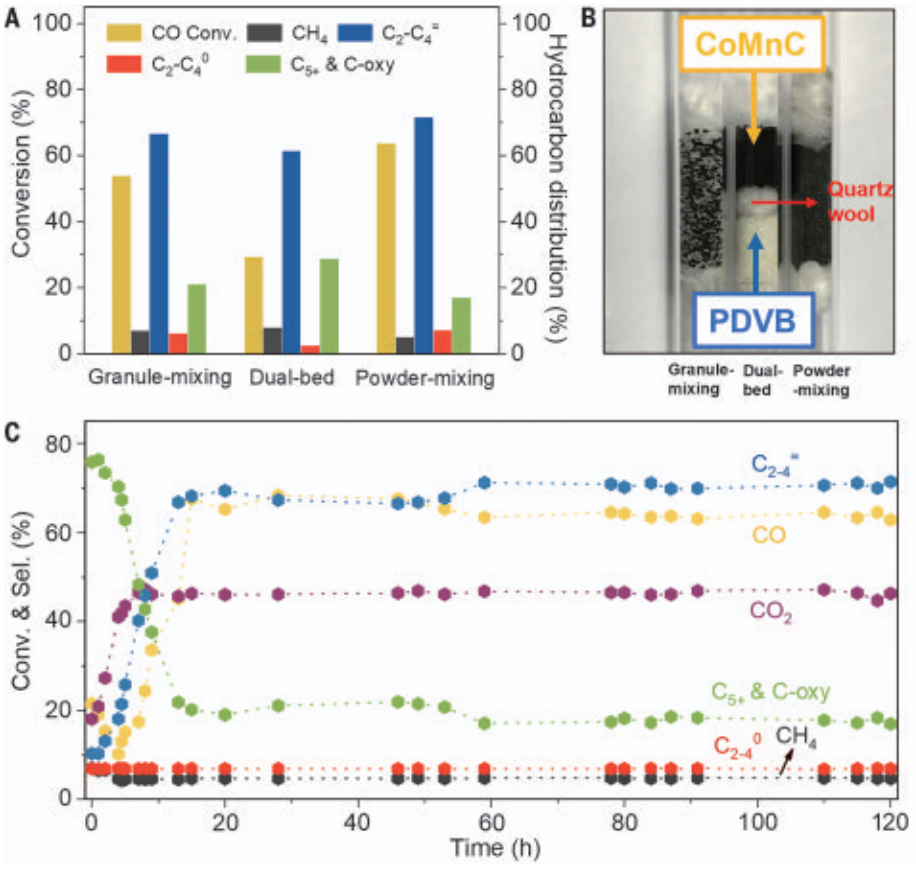


Fig. 2. Effect of different physical mixture methods and catalyst durability. (**A**) Syngas conversion performance of CoMnC/PDVB produced using different mixing techniques. Reaction conditions: 1.0 g of CoMnC and 1.0 g of PDVB, H_2/CO at 2, 1800 ml g_{CoMnC}^{-1} $hour^{-1}$, 0.1 MPa, 250°C. (**B**) Photographs of the catalyst beds with different mixing techniques. (**C**) Durability of CoMnC/PVDB catalyst in the syngas conversion to light olefins. Reaction conditions: 1.0 g of CoMnC catalyst physically mixed with 1.0 g of PDVB (powder-mixing), H_2/CO at 2, 1800 ml g_{CoMnC}^{-1} $hour^{-1}$, 0.1 MPa, 250°C.

powder mixture of the CoMnC and PDVB, the granule mixture, which was prepared by granulating the CoMnC and PDVB components and then mixing them together (40 to 60 mesh), showed lower CO conversion of 53.7% and similar C_2-C_4 olefin selectivity (66.4%). In a dual-bed reactor, the PDVB was packed below the CoMnC catalyst bed and separated by a layer of inert quartz sand. The CO conversion of 29.1% and lower olefin selectivity of 61.2% were similar to those of bare CoMnC catalyst. This result indicated that PDVB is inert for the reaction and that the promotion with PDVB required physical mixing with the CoMnC catalyst.

The CoMnC/PDVB catalyst was used in a continuous reaction test to evaluate durability. After activation for ~15 hours, the CO conversion was constant at steady state with an average value of ~64.7% (Fig. 2C). Even after reaction for 120 hours, the CO conversion was well maintained at 62.8% with stable CoMnC and PDVB components, confirming the good durability of the CoMnC/PDVB. In this process, the selectivity of light olefins was also constant at 70.0%, with an average productivity at 5.9 mmol g_{CoMnC}^{-1} $hour^{-1}$. During the test, the

selectivities for undesired methane and C_2–C_4 alkanes remained lower than 5.0% and 7.0%, respectively. The selectivity for C_{5+} products was ~18.0%, and 92.0% of those are C_5–C_8 olefins (*19*, *20*).

To understand the promotion of PDVB in the syngas conversion, we used temperature-programmed surface reaction mass spectrometry (TPSR-MS) by feeding syngas to the catalysts. The CoMnC catalyst has been reported to have superior activity for hydrogen activation and cleavage of the C-O bond (*8*, *21*–*23*), and these steps can be identified in TPSR-MS tests. The dependences of propylene, methane, and water signals (*m/z* = 42, 16, and 18, respectively) on reaction temperatures (Fig. 3A) show that on the CoMnC catalyst, the water and methane signals appeared at 166° and 203°C, which we assigned to the C-O cleavage and hydrogenation reaction. At 217°C, the propylene signal began to appear because the C-C coupling step occurred after C-O dissociation. Similar signals also appeared on the CoMnC/PDVB catalysts, but the signals of propylene and water were obviously stronger than those on the bare CoMnC, indicating that the physically mixed PDVB indeed boosted the activity.

It might be expected that the PDVB could participate in the CoMnC carbonization that leads to the distinguishable catalytic performances (*22*, *24*). We excluded this hypothesis by characterizing the CoMnC phase change as a function of reaction time, which showed negligible difference in the x-ray diffraction (XRD) patterns of the CoMnC with and without PDVB during the reaction periods (fig. S7); this result was also supported by the TEM characterization (fig. S8). After removal of the PDVB component from the used CoMnC/PDVB catalyst, the resulting CoMnC component exhibited performance comparable to that of the as-prepared CoMnC (table S1). These results further indicate that the physical mixture with PDVB did not change the catalyst structure, and they are consistent with the high stability of PDVB at the reaction temperature (figs. S9 to S12).

The previous strategies of chemically modifying the catalyst surface with hydrophobic organosilanes were developed to improve water resistance in syngas conversion (*25*–*27*). Following this route, we also modified the CoMn catalyst using tetraethyl orthosilicate and dimethyl diethyloxysilane, but this resulted in lower CO conversion over the CoMn@Si and CoMn@Si-*c* catalysts relative to the bare CoMnC catalyst (table S1). XRD patterns of the used catalysts showed the presence of oxide phases and suggested that hindered carbonization led to formation of the active CoMn carbides (fig. S13) (*27*).

We also varied the amount of PDVB in the catalyst bed. PDVB/CoMnC weight ratios of 0.5, 1.0, and 1.5 resulted in distinguishable CO conversions at 56.4%, 63.5%, and 70.5%, respectively (table S7). The sensitivity of CO conversion to PDVB content indicated that it plays a crucial role in catalysis. Although PDVB in the physical mixture did not change the catalyst structure, it might optimize the water diffusion because of its hydrophobicity. Thus, we studied the role of added water in CO conversion over the CoMnC catalyst (Fig. 3B). Considering that the water production rate was 8.8 to 23.5 mg g_{CoMn}^{-1} $hour^{-1}$ in the CoMnC/PDVB-catalyzed syngas conversion (calculated according to the oxygen balance in the reaction system and the amount of collected water product after reaction; fig. S14), we added water at this rate to investigate its influence on the CO conversion. The CoMnC catalyst showed CO conversion of 33.5% at the beginning of the reaction without water injection, which then decreased to ~8.2% (average value) after water injection with a feed rate of ~10.5 mg g_{CoMn}^{-1} $hour^{-1}$. Interestingly, addition of PDVB to the catalyst efficiently minimized the negative effect of water, which exhibited only a relatively slight decrease in CO conversion to ~57.4% under the equivalent water feed. The activity of the CoMnC catalyst was continuously reduced as more water was added to the feed gas (fig. S15), and the CO conversion dropped to 4.7% with water feed rate of ~20.0 mg g_{CoMn}^{-1} $hour^{-1}$. These data confirmed the water-restricted feature of the CoMnC-catalyzed syngas conversion. In contrast, the profile of CO conversion as a function of water concentration was flatter over the CoMnC/PDVB catalyst. This hydrophobic material may have helped the water product to rapidly desorb after it was formed on the CoMnC surface and also hindered its readsorption (fig. S16), which would free up active sites for the continuous conversion of more CO molecules.

This hypothesis was supported by a pulse experiment to explore the CO adsorption on the catalyst surface with and without water injection (Fig. 3C). The CoMnC and CoMnC/PDVB catalysts were localized within the flowing He atmosphere at 250°C, with periodic pulsing of the CO or mixture of CO and water. In the test without water, the CO signals on

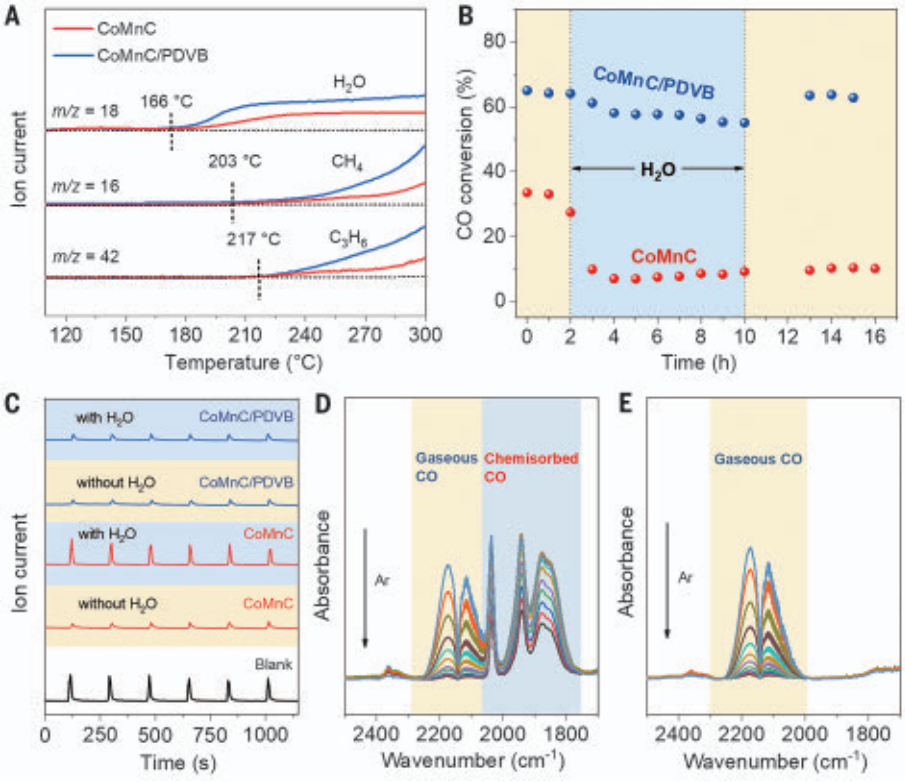


Fig. 3. PDVB-optimized water sorption. (**A**) The dependences of water, methane, and propylene signals (*m/z* at 18, 16, and 42) on temperature-programmed surface reaction by feeding syngas to the CoMnC and CoMnC/PDVB catalysts. (**B**) Data showing the influence of water on syngas conversion with CoMnC and CoMnC/PDVB. Reaction conditions: 1.0 g of CoMnC or 1.0 g of CoMnC physically mixed with 1.0 g of PDVB (powder-mixing), H_2/CO at 2, 1800 ml g_{CoMnC}^{-1} $hour^{-1}$, 0.1 MPa, 250°C. The water feed rate was ~10.5 mg g_{CoMn}^{-1} $hour^{-1}$. (**C**) Transient response curves obtained during pulses of 10% CO/He (5 ml/min) into pure He flow (30 ml/min) at 250°C over the CoMnC and CoMnC/PDVB catalysts. CO flowed through the water at 40°C for introducing water to the catalysts. (**D** and **E**) CO desorption in situ FTIR spectra of anhydrous CoMnC (D) and water-pretreated CoMnC catalysts (E).

both catalysts were extremely weak compared with that in the blank run without catalysts, indicating that efficient CO adsorption occurred on both CoMnC and CoMnC/PDVB catalysts. In these cases, the CO signals were similar, revealing the negligible effect of PDVB on CO sorption, and the CO adsorption dominantly occurred on the CoMnC. When water was fed into the reaction (~7 vol% in CO), the CO pulse peaks on the CoMnC catalyst were markedly increased, with the intensities similar to that in the blank run, revealing the hindered CO adsorption by competition with water.

Under the equivalent test, the CO pulse peaks were still weak on the CoMnC/PDVB catalyst, similar to the result of the water-free test, confirming the negligible influence of water on CO sorption in the presence of PDVB. A similar phenomenon was observed in the equivalent tests by changing the water feed amount (fig. S17). On the basis of these results, we conclude that the water on the catalyst surface could hinder the CO adsorption, while the PDVB efficiently shifted the sorption equilibrium described by $*H_2O \rightleftharpoons * + H_2O$ by accelerating the rapid water desorption and hindering its readsorption. These features led to a higher proportion of free catalyst surface for continuous conversion of CO, boosting the syngas conversion to olefins (fig. S18).

In situ Fourier transform infrared (FTIR) spectra characterizing the CO hydrogenation on the CoMnC and CoMnC/PDVB catalysts are shown in fig. S19. Introducing CO and hydrogen to the CoMnC catalyst led to the formation of obvious bands at 2717 to 2955 cm^{-1}, 1320 to 1527 cm^{-1}, and ~1039 cm^{-1}, which were assigned to the olefin species from CO hydrogenation and C-C coupling (*28*–*30*). The broad signals at ~1600 cm^{-1} and 3400 to 3600 cm^{-1} appeared and continuously increased with reaction time because of the water product adsorbed on the catalyst surface (*31*–*34*). On the CoMnC/PDVB catalyst, the signals of olefin products were solely observed with an extremely weak water signal, suggesting the rapid desorption of water once it is formed on the catalyst.

The effect of water on CO adsorption was further explored (Fig. 3, D and E). Introducing CO to the CoMnC catalyst led to signals of chemically adsorbed CO on the CoMnC surface (*22*, *33*). When water was co-fed with CO, the signals of chemically adsorbed CO were almost undetectable with only the bands of gaseous CO (*34*), confirming that CO adsorption was hindered under competition with water. This phenomenon might explain the negative effect of water on the syngas conversion by the competitive adsorption, in good agreement with the results of pulse experiments. In contrast, the catalyst containing PDVB exhibited comparable CO adsorption with and without water injection in a FTIR

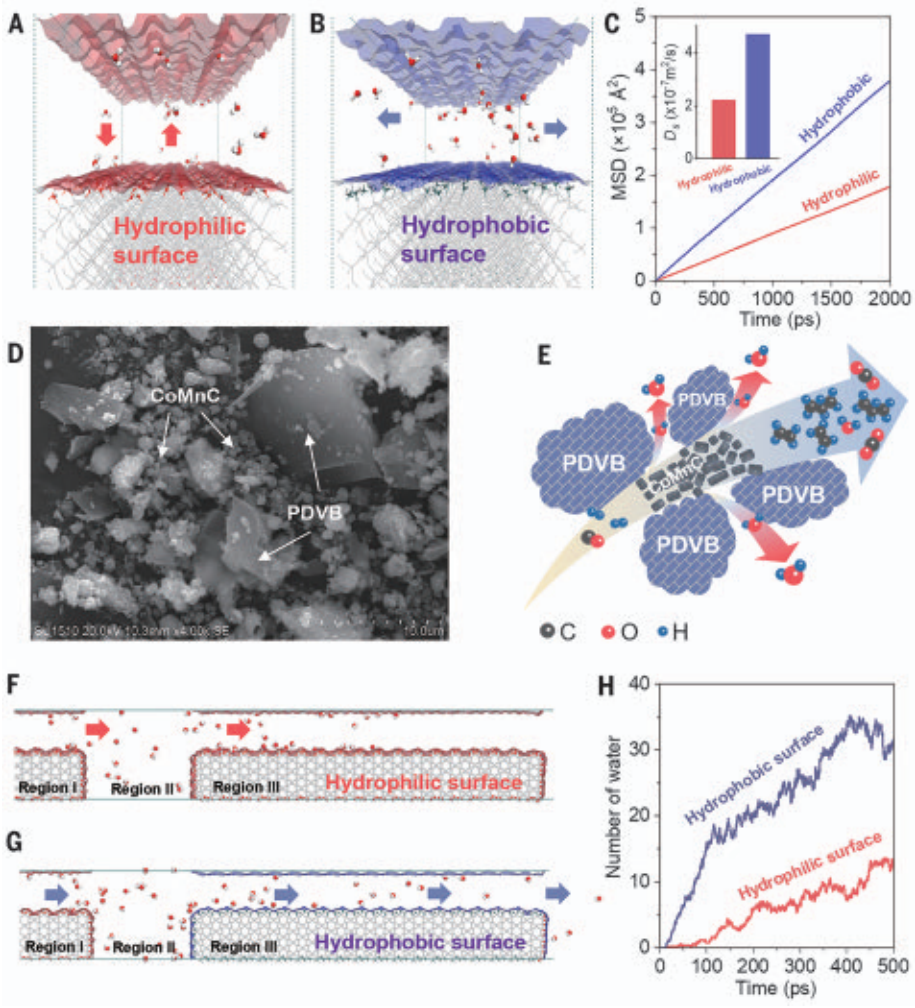


Fig. 4. Theoretical simulation. (**A** and **B**) Models showing the water diffusion within regions surrounded by hydrophilic and hydrophobic surfaces. (**C**) Mean square displacement (MSD) and diffusion coefficient (D_s) showing the water diffusion efficiency at 250°C. (**D**) SEM image of the CoMnC/PDVB granule. (**E**) Scheme showing the escape of water from the CoMnC surface through the region surrounded by PDVB. (**F** and **G**) Models showing the water escape through different regions. The regions I are hydrophilic and the regions III are hydrophilic and hydrophobic, respectively. (**H**) The number of water molecules that escaped from region III from an initial state with 100 water molecules on region I as a function of time.

study (figs. S20 and S21). According to previous studies, less water on the catalyst surface might reduce the concentration of carbon-based intermediates and enhance the hydrogenation activity (*35*, *36*), which could explain the improved selectivity of short hydrocarbons and slightly reduced *o*/*p* ratios on the CoMnC/PDVB catalyst relative to bare CoMnC.

In addition, the olefin products adsorbing on the catalyst surface might also hinder the CO conversion to some extent, as confirmed by catalysis studies with ethylene in the syngas feed (fig. S22). However, further tests suggested a negligible effect of PDVB on olefin sorption (fig. S23). Thus, PDVB accelerated syngas conversion in the catalytic tests; this was primarily attributed to its hydrophobicity in removing water rather than olefins. We further investigated the influence of promoter wettability on water diffusion by a theoretical simulation, in which we explored water molecule diffusion in a region surrounded by hydrophilic or hydrophobic surfaces (Fig. 4, A and B). Water diffusion efficiency was quantified by the diffusion coefficient D_s. The hydrophilic surface interacted with water molecules to slow down the transportation (Fig. 4C), giving $D_s = 2.2 \times 10^{-7}$ m^2/s. In the region surrounded by hydrophobic surfaces, the water diffusion was accelerated with obviously higher $D_s = 4.7 \times 10^{-7}$ m^2/s (Fig. 4C), given the relatively weak interaction between water and the hydrophobic surface (*37*). The theoretical simulation therefore gives a qualitative trend of the diffusion of water molecules along the different surface.

In our work, the CoMnC catalyst and PDVB were physically mixed and randomly distributed

in the catalyst granules. As observed in the scanning electron microscope (SEM) image of an actual CoMnC/PDVB catalyst granule (Fig. 4D), the CoMnC and PDVB were packed tightly and disordered, with intergranular distances ranging from nanometers to micrometers (Fig. 4E). We studied how the hydrophobic promoter affects the escape of water that is produced on the relatively hydrophilic surface (e.g., the CoMnC surface). In Fig. 4, F and G, we show models simulating the water escape from the solid surface (region I) through the region surrounded by hydrophobic or hydrophilic surfaces (region III), respectively. Notably, these regions were separated from each other (region II) for simulating the physical mixing with intergranular distances between the CoMnC and PDVB granules in the catalyst (figs. S24 and S25).

The number of escaped water molecules as a function of diffusion time in the systems with hydrophobic and hydrophilic promoters, from an initial state with 100 water molecules on region I (Fig. 4H), indicated that more water molecules escaped through the hydrophobic channel than from the hydrophilic channel under the equivalent conditions. For example, after 500 ps, ~32% of the initial water molecules escaped from the model with the hydrophobic channel, whereas only 13% escaped from the model with the hydrophilic channel. The influence of water concentration on diffusion rate was also simulated by regulating the number of water molecules in region I of the initial state (e.g., 25, 50, and 100 water molecules; fig. S26). The results showed that increasing the concentration of water molecules on the hydrophobic model surface could accelerate the diffusion rate of water molecules, which helps to explain the rapid water diffusion in the syngas conversion reaction with continuously produced water molecules. The models showed that the hydrophobic promoter physically regulated the catalyst by accelerating the water diffusion, in good agreement with the experimental results.

PDVB-promoted water sorption can also be directly observed through a model experiment of $CuSO_4{\cdot}5H_2O$ dehydration, because of its color change from blue to white upon dehydration. After mixing a small amount of PDVB to the $CuSO_4{\cdot}5H_2O$ (0.75 wt% of PDVB in the mixture), the color change was obviously accelerated, as confirmed by the photographs in fig. S27. This result again confirms that PDVB promoted water desorption and hindered readsorption when physically mixed.

We prepared physical mixtures of CoMnC catalyst with different materials whose wettability was distinguishable. When nanopores were introduced to the PDVB (two nanoporous PDVB materials with distinguishable surface areas at 488.2 and 623.3 m^2/g; table S2 and figs. S28 to S31), the CO conversion further improved to 88.7% and 92.4% over the CoMnC/nanoporous PDVB catalysts, but selectivity for light olefins was at 53.4% and 37.6% (table S1). Relative to CoMnC/PDVB, the CoMnC/nanoporous PDVB catalysts showed higher selectivity for the heavier olefins (C_5 to C_8 selectivity of 33.7% and 43.7%, respectively). This phenomenon might be the result of the high adsorption capacity of nanoporous PDVB for the olefin products (fig. S32), which prolonged the retention time of olefin products in the catalyst bed to benefit chain growth. These data confirmed that product distribution could be adjusted by changing the nanoporosity of the PDVB promoter.

Similar trends were also observed in the reaction with the CoMnC catalyst mixed with methyl group–modified silica with a hydrophobic surface (SiO_2-Me, figs. S33 to S35), which showed CO conversion at 74.7% and lower selectivity for C_2–C_4 olefins at 45.5% (table S1). Higher selectivity for C_5–C_8 products and lower *o/p* ratios were obtained than with the reaction over CoMnC/PDVB. When graphite, a hydrophobic carbon material, was mixed with CoMnC (CoMnC/Gra, fig. S36), the CO conversion was 53.0% with 67.3% selectivity for light olefins (table S1). Given that the graphite is earth-abundant and extremely cheap, our strategy for shifting water-mediated sorption equilibrium could be implemented simply by mixing hydrophobic graphite with the current catalysts.

In addition, when relatively hydrophilic materials (such as a mixture of PDVB, SiO_2, and hydrophilic polymers; figs. S37 to S40) were used in the CoMnC-catalyzed syngas conversion, the CO conversion was markedly reduced (tables S8 to S10). For example, polystyrene (PS, fig. S40), which has composition similar to PDVB but is more hydrophilic, failed to promote the CoMnC-catalyzed syngas conversion, showing CO conversion at 10.7% with methane selectivity of 37.8% and C_2–C_4 olefin selectivity of 37.8% (table S10). These data show the importance of a hydrophobic promoter.

Our approach could be used to upgrade industrially catalytic processes without modifying the catalysts themselves. In addition, the strategy is conceptually different from catalyst hydrophilization with organosilanes, described as a chemical modification route, where the conversion was not obviously improved in syngas conversion but the CO_2 selectivity was reduced by hindering the undesired water-gas shift (*4*). This difference might result from the distinguishable distances between the active site and the hydrophobic surface for these different systems. Considering that many hydrogenation reactions are strongly affected by water, the physical regulation method using a promoter with desired wettability could guide the design of more efficient catalysts in the future.

REFERENCES AND NOTES

1. M. Wolf, N. Fischer, M. Claeys, *Nat. Catal.* **3**, 962–965 (2020).
2. H. Li *et al*., *Science* **367**, 667–671 (2020).
3. S. M. Hashim, A. R. Mohamed, S. Bhatia, *Rev. Chem. Eng.* **27**, 157–178 (2011).
4. Y. Xu *et al*., *Science* **371**, 610–613 (2021).
5. S. Crossley, J. Faria, M. Shen, D. E. Resasco, *Science* **327**, 68–72 (2010).
6. H. Yang, T. Zhou, W. Zhang, *Angew. Chem. Int. Ed.* **52**, 7455–7459 (2013).
7. Z. Jin *et al*., *Science* **367**, 193–197 (2020).
8. L. Zhong *et al*., *Nature* **538**, 84–87 (2016).
9. R. Yang *et al*., *J. Energy Chem.* **32**, 118–123 (2019).
10. Y. Zhang *et al*., *Nano Today* **4**, 135–142 (2009).
11. P. Ramírez de la Piscina, J. L. G. Fierro, G. Muller, J. Sales, N. Homs, *Catal. Lett.* **14**, 45–49 (1992).
12. W. Fan, P. Wu, T. Tatsumi, *J. Catal.* **256**, 62–73 (2008).
13. R. Yang, R. Gao, Z. Qian, Y. Wang, *Separ. Purif. Tech.* **230**, 115884 (2020).
14. H. M. Torres Galvis *et al*., *Science* **335**, 835–838 (2012).
15. F. Jiao *et al*., *Science* **351**, 1065–1068 (2016).
16. K. Cheng *et al*., *Angew. Chem. Int. Ed.* **55**, 4725–4728 (2016).
17. S. Soled, E. Iglesia, R. A. Fiato, *Catal. Lett.* **7**, 271–280 (1990).
18. M. D. Shroff *et al*., *J. Catal.* **156**, 185–207 (1995).
19. T. Ren, M. Patel, K. Blok, *Energy* **31**, 425–451 (2006).
20. R. Snel, *Catal. Rev., Sci. Eng.* **29**, 361–445 (1987).
21. H. Wang *et al*., *J. Am. Chem. Soc.* **135**, 4149–4158 (2013).
22. Z. Li *et al*., *ACS Catal.* **7**, 3622–3631 (2017).
23. F. Jiang *et al*., *Catal. Today* **369**, 158–166 (2021).
24. Y. Yang *et al*., *Appl. Catal. A* **549**, 179–187 (2018).
25. R. Xie *et al*., *Catal. Commun.* **12**, 589–592 (2011).
26. M. Ojeda *et al*., *Langmuir* **22**, 3131–3137 (2006).
27. X. Yu *et al*., *Appl. Catal. B* **232**, 420–428 (2018).
28. C. J. Weststrate, J. W. Niemantsverdriet, *ACS Catal.* **8**, 10826–10835 (2018).
29. C. J. Weststrate, I. M. Ciobîcă, J. van de Loosdrecht, J. W. Niemantsverdriet, *J. Phys. Chem. C* **120**, 29210–29224 (2016).
30. A. A. Efremov, A. A. Davydov, *React. Kinet. Catal. Lett.* **15**, 327–331 (1980).
31. L. Marchese, J. Chen, J. M. Thomas, S. Coluccia, A. Zecchina, *J. Phys. Chem.* **98**, 13350–13356 (1994).
32. K. Tanaka, J. M. White, *J. Phys. Chem.* **86**, 4708–4714 (1982).
33. D. J. C. Yates, *J. Phys. Chem.* **65**, 746–753 (1961).
34. C.-J. Jia *et al*., *J. Am. Chem. Soc.* **133**, 11279–11288 (2011).
35. A. K. Dalai, B. H. Davis, *Appl. Catal. A* **348**, 1–15 (2008).
36. C. J. Bertole, C. A. Mims, G. Kiss, *J. Catal.* **210**, 84–96 (2002).
37. K. Liu, X. Yao, L. Jiang, *Chem. Soc. Rev.* **39**, 3240–3255 (2010).

ACKNOWLEDGMENTS

We thank F. Chen for help in SEM characterization. **Funding:** Supported by the National Key Research and Development Program of China (2021YFA1500404), National Natural Science Foundation of China (U21B20101, 21932006, 22032005, 22102143, and 22125304), China Postdoctoral Science Foundation (2021M702803), and National Postdoctoral Program for Innovative Talents (BX20200291). **Author contributions:** W.F. and C.W. performed the catalyst preparation, characterization, catalytic tests, and data analysis. Z.Q.L. and A.Z. performed the theorical calculation and wrote the corresponding part. L.L., H.L., X.Q., S.X., and L.J.L. participated in the catalyst characterization and discussion. L.W. and F.-S.X. designed the study, analyzed the data, and wrote the paper. **Competing interests:** The authors declare that there is no conflict of interest. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo0356
Materials and Methods
Figs. S1 to S40
Tables S1 to S10
References (*38*–*57*)

Submitted 9 January 2022; resubmitted 18 April 2022
Accepted 24 May 2022
10.1126/science.abo0356

ORGANIC CHEMISTRY

A concise synthesis of tetrodotoxin

David B. Konrad[1]†‡, Klaus-Peter Rühmann[2]‡, Hiroyasu Ando[3], Belinda E. Hetzler[2], Nina Strassner[4], Kendall N. Houk[4], Bryan S. Matsuura[2]*‡, Dirk Trauner[2]*§

Tetrodotoxin (TTX) is a neurotoxic natural product that is an indispensable probe in neuroscience, a biosynthetic and ecological enigma, and a celebrated target of synthetic chemistry. Here, we present a stereoselective synthesis of TTX that proceeds in 22 steps from a glucose derivative. The central cyclohexane ring of TTX and its α-tertiary amine moiety were established by the intramolecular 1,3-dipolar cycloaddition of a nitrile oxide, followed by alkynyl addition to the resultant isoxazoline. A ruthenium-catalyzed hydroxylactonization set the stage for the formation of the dioxa-adamantane core. Installation of the guanidine, oxidation of a primary alcohol, and a late-stage epimerization gave a mixture of TTX and anhydro-TTX. This synthetic approach could give ready access to biologically active derivatives.

Tetrodotoxin (TTX) is a neurotoxic natural product that has inspired and empowered chemists and biologists for more than a century (*1*–*3*). As a selective blocker of voltage-gated sodium channels, it has played a crucial role in the elucidation of the action potential, and it is still routinely used to silence excitable cells in neural systems. Its isolation from widely differing species, such as pufferfish, starfish, sea snails, octopi, toads, and newts, has prompted intense investigations into its true biological producers, its biosynthesis, and its ecological role. It is now clear that TTX is synthesized by bacteria and accumulated by metazoan hosts as a defense against predators (*4*). Its toxicology and therapeutic utility in humans have been studied for decades and are still a topic of ongoing research (*5*).

As a synthetic target, TTX has been celebrated for the sheer intellectual challenge it provides and for the opportunity to demonstrate methodological and strategic advances. Its simple carbon framework, consisting of a cyclohexane ring with C1 and C2 side chains, stands in stark contrast to the dense network of polar functional groups that adorn it. Two hydroxy groups in a *syn* relationship engage a carboxylate as an ortho acid to form the signature dioxa-adamantane core of TTX, which is fused to a cyclic guanidine via an α-tertiary amine. One primary, two secondary, and a tertiary hydroxy group, as well as a hemiaminal, contribute further to the structural complexity of the molecule, which features four rings and nine contiguous stereocenters.

[1]Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstr. 5-13, 81377 Munich, Germany. [2]Department of Chemistry, New York University, 100 Washington Square East, New York, NY 10003, USA. [3]Graduate School of Infection Control Sciences and Kitasato Institute for Life Sciences, Kitasato University, Tokyo, Japan. [4]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA. ***Corresponding author. Email: b2matsuura@gmail.com (B.S.M.); dtrauner@upenn.edu (D.T.)**
†Present address: Department of Pharmacy, Ludwig-Maximilians-Universität München, Butenandtstr. 5-13, 81377 Munich, Germany.
‡These authors contributed equally to this work.
§Present address: Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104-6323, USA.

The first total synthesis of TTX, in racemic form, by Kishi and Fukuyama in 1972 stands as a landmark achievement in organic synthesis that, at the time, seemed hard to surpass (*6*). After a pause of more than 30 years, Isobe and co-workers published the first asymmetric synthesis in 2003 (*7*). This was followed shortly by the Hinman and Du Bois' asymmetric synthesis (2003) (*8*), a second Isobe approach (2004) (*9*, *10*), and a racemic and two asymmetric syntheses by Sato's group (2005, 2008 and 2010, respectively) (*11*–*13*). In 2017 and 2020, Fukuyama, Yokoshima, and colleagues revisited the molecule and published two distinct asymmetric routes to TTX (*14*, *15*). In addition, several studies have been published that intercept late-stage intermediates of the previous syntheses e.g., by the Alonso (2010) (*16*, *17*), Ciufolini (2015) (*18*–*20*), and Hudlicky (2018) (*21*) groups. Other approaches toward the molecule have been outlined (*3*, *22*, *23*).

An analysis of previous syntheses revealed several common features that motivated us to pursue a distinct strategy: (i) The cyclohexane core was either incorporated in the starting material, or was formed early, and then oxygens were added using epoxidations, dihydroxylations, or allylic oxidations of strategically placed alkenes. (ii) The α-tertiary amine on C8a was established via C–N bond formation, which must overcome considerable steric hindrance. Several methods have been implemented to address this challenge, such as intramolecular nitrogen transfer (sigmatropic rearrangement, aza-conjugate addition, nitrene insertion) and intermolecular S_N1-type nucleophilic substitution. (iii) The dioxa-adamantane was always formed spontaneously with careful orchestration of the sequence to adjust the oxidation state of C10 and set the labile C9 stereocenter. (iv) Every synthesis introduces the guanidine at a late stage (*7*, *24*) and uses protecting groups amenable to global deprotection in the final step.

Our synthetic analysis was guided by an attempt to link the formation of the cyclohexane core with the establishment of the α-tertiary amine as closely as possible, in contrast to previous total syntheses in which these strategic key steps were largely independent (*17*). To this end, we established a linear precursor that contained all the oxygen functionalities of the TTX skeleton, which would then be conjoined by a ring-forming reaction. This would be followed by installation of the α-tertiary amine through C–C bond formation, to introduce the C2 fragment that would eventually be incorporated into the dioxa-adamantane. Finally, we needed to develop a method to oxidize and lactonize this C2 fragment in a highly efficient manner.

Our ultimate retrosynthetic analysis is summarized in Fig. 1. Although not all compounds shown therein were defined in such detail at the outset of our study, it captures the essence of our synthetic plan. We reasoned that we could trace TTX back from an oxidation of an alkynyl isoxazolidine of type **1**, which would stem from bicyclic isoxazoline **2**, the product of an intramolecular 1,3-dipolar cycloaddition. Nitromethane would serve as a key linchpin in the assembly of **2**, reacting first in an intermolecular Henry reaction with aldehyde ***3***, followed by a dehydration to generate a reactive nitrile oxide intermediate that would close the central cyclohexane ring within a (3+2) cycloaddition. We have previously developed an asymmetric synthesis of unsaturated aldehydes similar to ***3*** via a Kiyooka aldol reaction and used it toward kweichowenol A, a polyoxygenated cyclohexene isolated from the plant *Uvaria kweichowensis* (*25*). Although an analogous route gave the aldehyde ***3*** in sufficient quantities to proceed with the synthesis of TTX, we found it more practical and economical to start from the glucose-derived building block **4**. All the carbons of glucose and two of its stereocenters would be retained over the course of the synthesis, making this an attractive starting material.

Previously known exo-methylene building block **5** was synthesized in three steps on a decagram scale from commercially available glucose derivative **4** and was also used by Sato and colleagues in their approach to TTX (see supplementary materials) (*13*). Regioselective reductive cleavage of the benzylidene acetal placed a benzyl ether at C5, yielding **6** (Fig. 2). A subsequent dihydroxylation then installed the tertiary alcohol with the correct absolute configuration at C6, as well as the C11 primary alcohol of TTX, providing **7** in excellent yield and as a single diastereomer (*26*). Protection of the vicinal diol as the acetonide, followed by an Appel reaction, gave the primary iodide **8**. Under conditions developed by Soengas and Silva, **8** underwent a reductive cleavage upon treatment with *tert*-butyl lithium at low temperature, to yield a δ,ε-unsaturated aldehyde (**12**, Fig. 3), which engaged in a Henry reaction in situ upon addition of nitromethane. This afforded nitro alcohols **9a**,**b** as a separable 1:1 mixture of diastereomers (*27*).

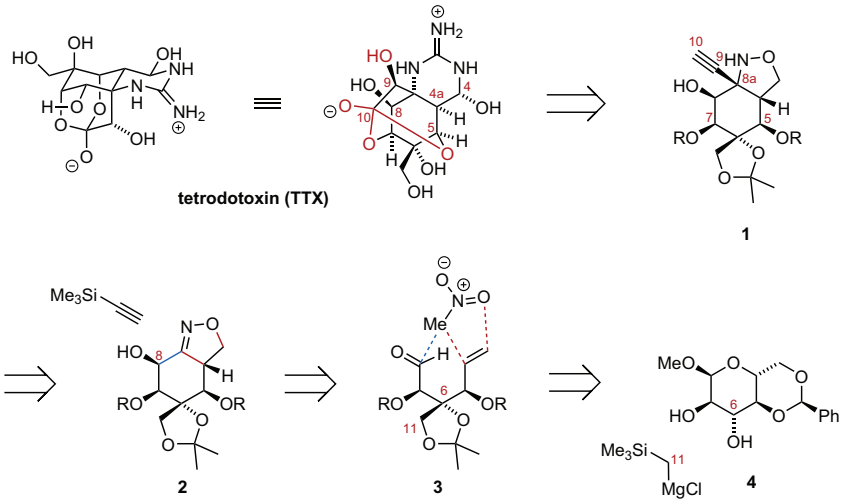


Fig. 1. Retrosynthetic analysis and synthetic design.

With nitro alcohols **9a**,**b** in hand, we were ready to attempt the key nitrile oxide cycloaddition to form the cyclohexane core of tetrodotoxin. Treatment of diastereomer **9a** with phenyl isocyanate and catalytic amounts of triethylamine triggered a dehydrative nitrile oxide cycloaddition, yielding isoxazoline **10** in moderate yield as a single diastereomer (*28*). An x-ray crystallographic structure of a byproduct, *N*-phenyl carbamate **11**, confirmed the configurations of the C8 and C4a stereocenters. Although the C4a stereocenter was inverted with respect to TTX, we reasoned that this isoxazoline diastereomer would be more accessible to nucleophiles for the formation of the hindered α-tertiary amine (*29*). Exposing diastereomer **9b** to the same reaction conditions only resulted in decomposition.

The poor diastereoselectivity of the Henry reaction and low yields of the cycloaddition severely hampered material throughput. Moving forward, we needed to develop a strategy that would selectively install the stereocenter at C8—ideally with a protected hydroxyl—which is critical in the ensuing cycloaddition. We reasoned that this could be accomplished via a dehydration to the corresponding nitroalkene followed by a conjugate addition with an appropriate *O*-nucleophile. This was realized by subjecting **8** to the same reductive fragmentation and Henry reaction sequence, followed by in situ dehydration, which yielded nitroalkene **14** exclusively as the *E*-isomer. **14** underwent an oxa-Michael addition upon treatment with the lithiated alkoxide of *p*-anisyl alcohol, presumably resulting in nitronate anion **15**. This intermediate could be intercepted with Boc-anhydride (via **16**) to trigger the formation of the transient nitrile oxide **17** and subsequent 1,3-dipolar cycloaddition, affording isoxazoline **18** in high yield and on a decagram scale (*30*). This reaction cascade was exquisitely diastereoselective, affording the central cyclohexane core of TTX with the correct configuration at C8, albeit with the wrong configuration at C4a.

Both the diastereoselectivity of the oxa-Michael addition and the mechanism of the 1,3-dipolar cycloaddition were further explored with quantum mechanical density functional theory calculations (see supplementary materials). The addition of PMB alkoxide, which resulted in the stereoselective formation of **15**, was shown to be favored by at least 3.4 kcal/mol (ΔG^{273K}) relative to the other five possible transition states. In the major transition state, the carbon chain adopts an *anti*-configuration with respect to the forming bond, as the Felkin-Anh model proposes, and the alkoxy group is *sym* to the double bond, an "inside alkoxy" arrangement shown to be favored for related cycloadditions (*31*).

The 1,3-dipolar cycloaddition can proceed through two possible pathways, initiated either by elimination of the *tert*-butyl carbonate (Boc-OH), followed by nitrile oxide cycloaddition, or by cycloaddition of the Boc-nitronate, followed by Boc-OH elimination. Calculations show that an elimination of Boc-OH proceeds quickly with a barrier of only 9.9 kcal/mol (ΔG^{298K}), whereas both cycloadditions occur with barriers of ~23 kcal/mol. Thus, the nitrile oxide pathway is strongly favored (see supplementary materials for details). The computational investigation further highlights a clear preference for the experimentally observed stereoisomer at C4a with a free energy difference of 1.4 kcal/mol. Although both transition states (shown in supplementary materials) adopt chair conformations, the one favoring the observed product avoids a 1,3-diaxial interaction of the alkoxy groups.

After deprotection of the *p*-methoxybenzyl (PMB) group with ceric ammonium nitrate (CAN), **10** was subjected to lithiated trimethylsilyl (TMS)–acetylide which, in the presence of $BF_3{\cdot}OEt_2$, underwent addition from the convex face of the bicyclic isoxazoline (*32*). After cleavage of the TMS group upon workup, isoxazolidine alkyne **19** was isolated as a single diastereomer, which possessed all the skeletal carbons of TTX and five out of its nine stereocenters with the correct absolute configuration. The presence of a free hydroxy group in **10**, which presumably coordinates to the nucleophile after deprotonation, was found to be crucial for the success of the addition to the oxime ether moiety. Isoxazoline **18**, by contrast, gave very low conversion. The lithiated TMS-acetylide was highly effective, whereas more functionalized C2 synthons were found to be unreactive.

Our next goal was the oxidative elaboration of the alkyne to introduce the stereogenic C9 alcohol and the C10 lactone that would ultimately engage in the formation of the signature ortho acid of TTX. Our original strategy was to form the requisite hydroxylactone via a gold- or silver-catalyzed 5-endo-dig hydroetherification of the C8 hydroxy group of **19** onto the alkyne terminus (Fig. 4A). Oxidation of the resultant dihydrofuran to a hydroxylactone, followed by isomerization, via a translactonization during the final deprotection, would unveil the central ortho acid. Unfortunately, extensive experimentation proved that this approach is exceptionally difficult to execute, requiring us to reexamine our sequence of bond-forming events. The logical solution to this conundrum would be to engage the alkyne terminus in a bridge-forming hydroetherification with either the C5 or C7 hydroxy group, followed by oxidation to the desired hydroxylactone, which would obviate the need for a translactonization step. However, this strategic shift was not without risk, as such bridge-forming 6-endo-dig cyclizations have scarce precedent.

To this end, we needed to establish a protecting group scheme that would also be compatible with the subsequent introduction of the guanidine moiety and the final steps. Exposure of **19** to Boc_2O gave **20**, which features both a *tert*-butyl carbonate and carbamate moiety. The next critical deprotection step required the simultaneous cleavage of the C5 and C7 benzyl ethers in the presence of the alkyne and isoxazolidine groups, which are sensitive to hydrogenation conditions, as well as three acid-sensitive protecting groups. After considerable experimentation, we found that the benzyl ethers could be cleanly removed by using Pieber and Seeberger's recently introduced chromoselective photochemical debenzylation (*33*), a singularly effective protocol that afforded diol **21** in excellent yield. The Boc-group on the C8-alcohol and the use of a 525-nm green light-emitting diode as the irradiation source were key to this reaction's success. The use of either the unprotected alcohol or higher-energy blue light resulted in complete substrate decomposition. Following the debenzylation, a selective methanolysis of the carbonate gave a triol, which could be protected to bis-acetonide **22**. At this stage, we elected to

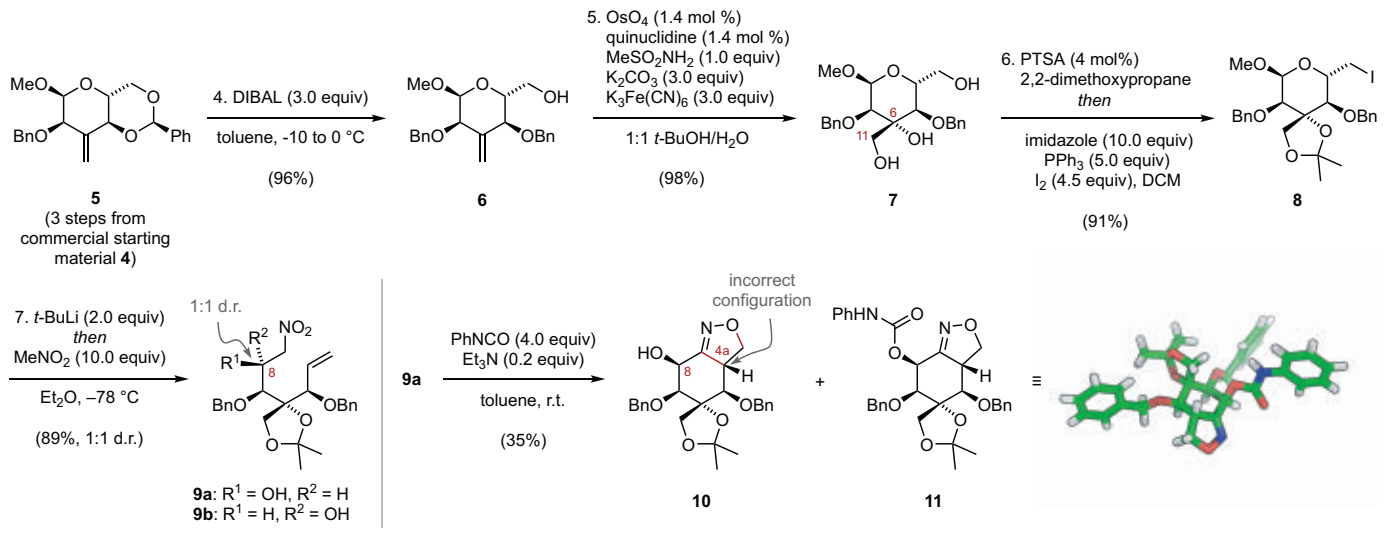


Fig. 2. Opening sequence and initial attempts to form the carbocyclic core. DIBAL, diisobutylaluminium hydride; PTSA, *p*-toluene sulfonic acid; DCM, dichloromethane; PhNCO, phenylisocyanate; r.t., room temperature.

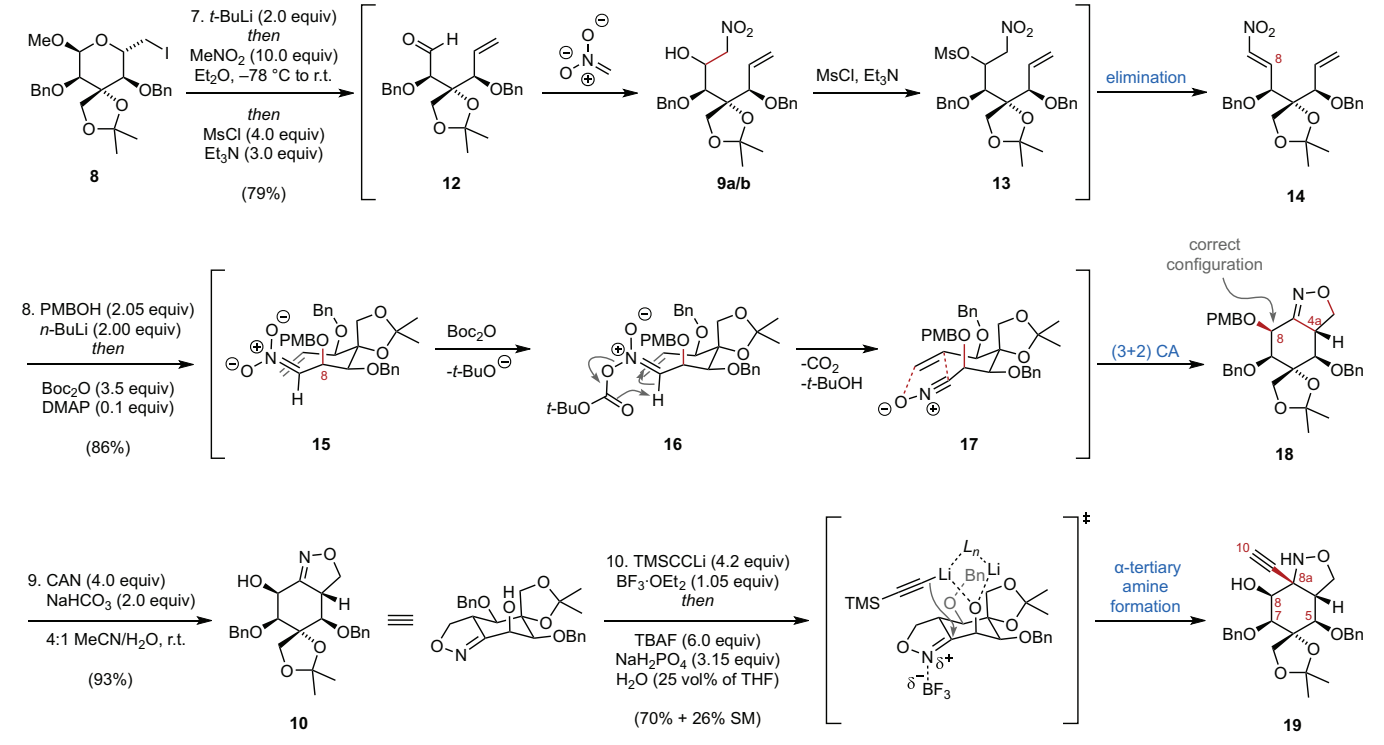


Fig. 3. Development of a diastereoselective route to the cyclohexane core and installation of the α-tertiary amine. MsCl, methanesulfonyl chloride; PMBOH, *p*-methoxybenzyl alcohol; Boc_2O, di-*tert*-butyl dicarbonate; DMAP, 4-dimethylaminopyridine; (3+2) CA, (3+2) cycloaddition; CAN, ceric ammonium nitrate; TMSCCLi, lithium trimethylsilylacetylide; TBAF, tetra-*n*-butylammonium fluoride; THF: tetrahydrofuran.

reduce the sensitive *N*-Boc isoxazolidine. By using SmI_2, the N–O bond was cleaved in excellent yield, resulting in a primary alcohol, which was then protected as its silyl ether **23** (Fig. 4A).

The stage was now set for the next key step of our synthesis, the conversion of alkyne **23** to hydroxylactone **25**. To take full advantage of the steric environment provided by the proximal acetonide, we aimed at forming the C10-O5 bond first to yield a dihydropyran that could then be oxidized to the hydroxylactone in a stereoselective manner. Initially, we attempted to achieve this via a 6-endo-dig cyclization with gold- or silver-based π-acid catalysts, but this approach was thwarted by the substrate's propensity to undergo an undesired 5-exo-dig cyclization.

Our solution to this problem was inspired by reports from Trost (*34*) and McDonald (*35*) on the catalytic generation of metallo-vinylidene carbenes, which would render the alkyne terminus (C10) electrophilic. We found that **23** could be converted to bridged dihydropyran **24**, with $CpRu(PPh_3)_2Cl$ as a cycloisomerization catalyst, in nearly quantitative yield on a 300-mg scale (*36*). Pushing this finding even further, we postulated that the resultant dihydropyran could be converted to the key hydroxylactone **25** by transforming the cycloisomerization catalyst into an oxidant. This hypothesis was supported by Blechert and co-workers, who demonstrated that ring-closing metathesis could be coupled with olefin dihydroxylation on simple bis-alkenes by using Ru-based metathesis catalysts (*37*). However, we would need to identify conditions that would further oxidize the hypothetical diol to the hydroxylactone, without overoxidation of

the desired product (e.g., oxidative cleavage or ketolactone formation). This was achieved by the addition of Oxone and a cosolvent mixture to the reaction (*38*). Presumably, under these conditions, the catalyst is oxidized to RuO_4, which, in turn, oxidized **24** to the desired hydroxylactone **25** with almost complete diastereoselectivity (Fig. 4B). We believe the chemoselectivity of the second oxidation is due to the increased hydricity of the C10–H bond of the hemiacetal intermediate. This single reaction combines the C10-O5 bond formation with two oxidation events, setting the C9 stereocenter in the processes, and had the added benefit of simplifying the purification, because the cycloisomerization catalyst was copolar with **24** on silica (*39*).

Having found a satisfying solution for the hydroxylactone problem, we decided to install

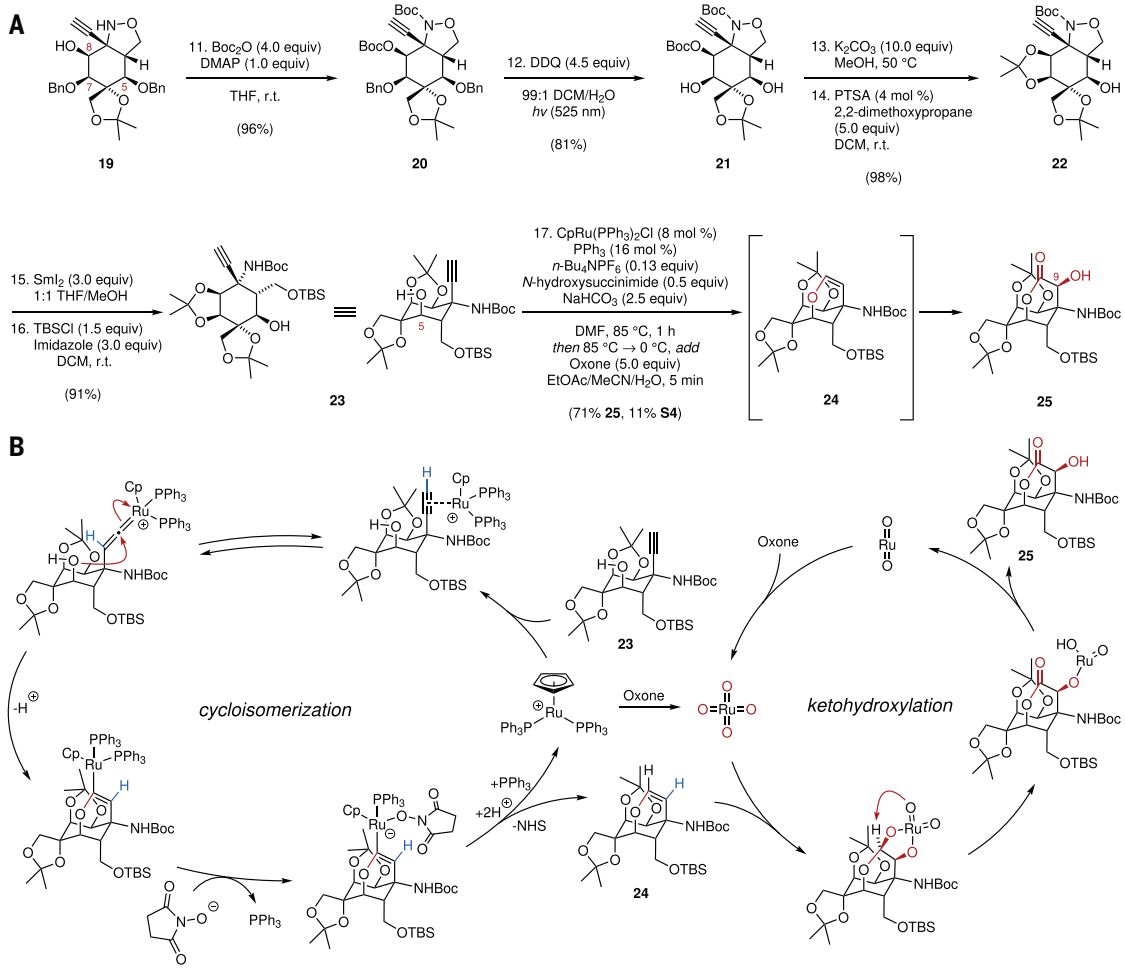


Fig. 4. Continuation of the synthesis and proposed mechanism for the key hydroxylactonization. (**A**) Synthesis. (**B**) Proposed mechanism. DDQ, 2,3-dichloro-5,6-dicyano-1,4-benzoquinone; TBSCl, *tert*-butyldimethylsilyl chloride; DMF, dimethylformamide; Cp, cyclopentadienyl.

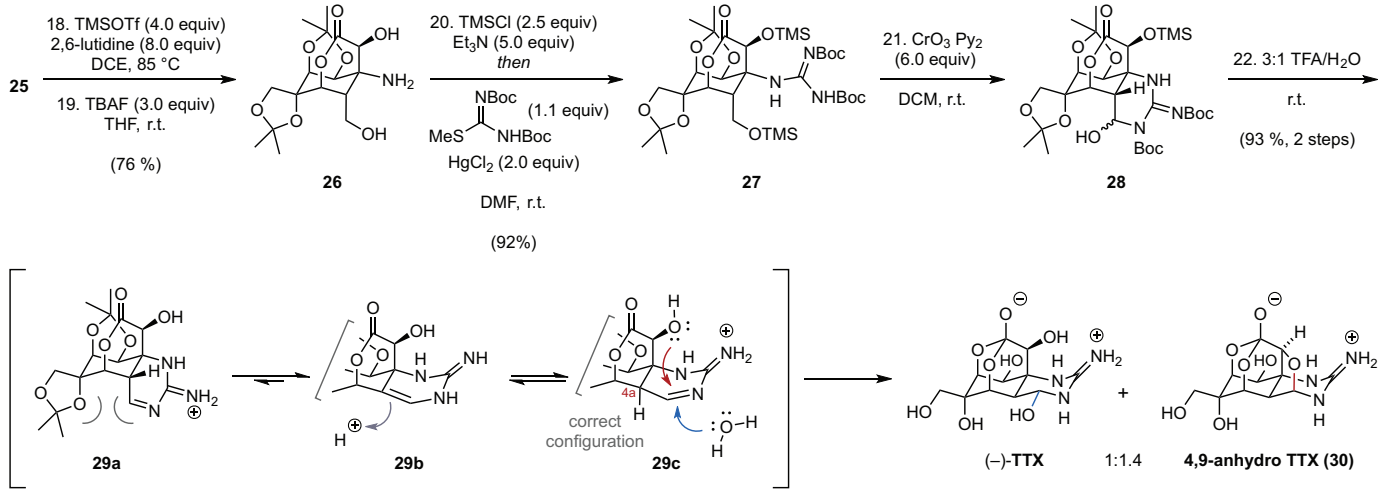


Fig. 5. Completion of the synthesis. TMSOTf, trimethylsilyl triflate; DCE, 1,2-dichloroethane; Py, pyridine; TFA, trifluoroacetic acid.

the guanidine moiety next. Removal of the *N*-Boc protecting group with trimethylsilyl triflate, followed by cleavage of the silyl ethers, one of which was transient, gave amino diol **26**. This compound was protected in situ as the bis-trimethyl silyl ether and then underwent clean guanylation under Kishi's conditions to afford compound **27**, which features all the atoms of TTX (Fig. 5).

In the last phase of our synthesis, we had to overcome two final obstacles: the oxidation of the primary silyl ether and the epimerization of the C4a stereocenter. A footnote in Kishi's seminal publication suggested the latter presented a potential liability owing to the propensity of the C5-O bond to undergo facile elimination, and all prior approaches had set this stereocenter earlier in their respective syntheses (*6*). Nevertheless, we continued by treating bis-trimethyl silyl ether **27** with Collins reagent, which effected selective deprotection and oxidation of the primary alcohol, resulting in **28** as a mixture of hemiaminal isomers. This transformation could not have been performed with a more stable *tert*-butyl dimethyl silyl ether in place, that is, with a guanidinylated derivative of compound **25**. Crude **28** was then dissolved in 25% aqueous trifluoroacetic acid and stirred at room temperature overnight. Notably, this effected the desired deprotections, epimerization, and cyclizations and gave a 1:1.4 mixture of TTX and 4,9-anhydro TTX (**30**) in good yield. We observed the formation of **30** in unusually high proportions, which could be explained by the guanidine participating in the epimerization process. Intramolecular condensation of the N3 nitrogen with the C4a aldehyde would form a highly stabilized iminium cation **29a**, which could undergo elimination to enamine **29b**. Protonation to the more thermodynamically favored iminium **29c** situates the electrophilic C4 in close proximity to the C9 hydroxy group, which can attack at a rate that is kinetically competitive with solvolysis. We believe that the elimination and tautomerization reactions are likely driven by unfavorable *syn*-pentane interactions between the C6 oxygen and the C4 iminium. However, given the multitude of transformations that are taking place in this final step, it is difficult, if not impossible, to pinpoint the exact sequence of events, which could take place in parallel and converge on TTX. TTX and **30** are known to be in equilibrium with one another, favoring TTX, and they can be readily interconverted (*7*, *40*). Indeed, upon heating of this mixture for 3 days as a solution in 5% d_3-AcOD–95% D_2O to 60°C, a 2.9:1 ratio of TTX and anhydro-TTX (**30**) was obtained (see supplementary materials). TTX and **30** have been separated on an analytical scale (*41*).

Taking stock of our strategic disconnections, our route showcases the power of the Huisgen cycloaddition for the construction of highly substituted cyclitols. For this purpose, the humble C1 synthon nitromethane performed spectacularly—involved in no fewer than six bond-forming and -breaking events—relaying an oxygen to C4, and embedding its nitrogen and carbon into the α-tertiary amine over the course of the synthesis. Although the bicyclic isoxazoline **10** possessed the incorrect stereocenter at C4a, necessitating a late-stage epimerization, it guided the subsequent acetylide addition from the convex face of the molecule to install the desired configuration of the C8a α-tertiary amine. Although this late-stage epimerization carried some strategic risk, we had reason to believe that it would succeed based on the coherence of the dioxa-adamantane core, which would render β-eliminations or retro-aldol reactions reversible. The alkynylation of **10** emphasizes the utility of oxime ethers as α-tertiary amine precursors but also highlights the need for more methodological development in this area. The Ru-catalyzed oxidative lactonization of alkyne **23** represents a notable advance in establishing the C9 and C10 hydroxylactone and should have future applications in the synthesis of other natural products such as the ginkolides and quassinoids (*42*, *43*). Furthermore, this synthesis served as a proving ground for the chromoselective photochemical debenzylation and recognizes the strategic value of implementing new technology and methods in highly complex settings.

Taken together, these strategic decisions resulted in one of the shortest and the most efficient syntheses of tetrodotoxin to date, accomplishing this goal in 22 total steps and 11% overall yield from commercially available starting materials. Our route is scalable and can be adapted to the production of other scarce tetrodotoxin derivatives to better understand their biosynthesis and chemical ecology. It is also amenable for the procurement of TTX derivatives that could serve as next-generation analgesics.

REFERENCES AND NOTES

1. J. Chau, M. A. Ciufolini, *Mar. Drugs* **9**, 2046–2074 (2011).
2. T. Nishikawa, M. Isobe, *Chem. Rec.* **13**, 286–302 (2013).
3. M. Makarova, L. Rycek, J. Hajicek, D. Baidilov, T. Hudlicky, *Angew. Chem. Int. Ed.* **58**, 18338–18387 (2019).
4. C. T. Hanifin, *Mar. Drugs* **8**, 577–593 (2010).
5. E. G. Moczydlowski, *Toxicon* **63**, 165–183 (2013).
6. Y. Kishi *et al*., *J. Am. Chem. Soc.* **94**, 9219–9221 (1972).
7. N. Ohyabu, T. Nishikawa, M. Isobe, *J. Am. Chem. Soc.* **125**, 8798–8805 (2003).
8. A. Hinman, J. Du Bois, *J. Am. Chem. Soc.* **125**, 11510–11511 (2003).
9. T. Nishikawa, D. Urabe, M. Isobe, *Angew. Chem. Int. Ed.* **43**, 4782–4785 (2004).
10. D. Urabe, T. Nishikawa, M. Isobe, *Chem. Asian J.* **1**, 125–135 (2006).
11. K. Sato *et al*., *J. Org. Chem.* **70**, 7496–7504 (2005).
12. K. Sato *et al*., *J. Org. Chem.* **73**, 1234–1242 (2008).
13. S. Akai *et al*., *Bull. Chem. Soc. Jpn.* **83**, 279–287 (2010).
14. T. Maehara, K. Motoyama, T. Toma, S. Yokoshima, T. Fukuyama, *Angew. Chem. Int. Ed.* **56**, 1549–1552 (2017).
15. K. Murakami, T. Toma, T. Fukuyama, S. Yokoshima, *Angew. Chem. Int. Ed.* **59**, 6253–6257 (2020).
16. F. Cagide-Fagín, R. Alonso, *Eur. J. Org. Chem.* **2010**, 6741–6747 (2010).
17. H. Lago-Santomé, R. Meana-Pañeda, R. Alonso, *J. Org. Chem.* **79**, 4300–4305 (2014).
18. B. A. Mendelsohn, M. A. Ciufolini, *Org. Lett.* **11**, 4736–4739 (2009).
19. J. Chau, S. Xu, M. A. Ciufolini, *J. Org. Chem.* **78**, 11901–11910 (2013).
20. S. Xu, M. A. Ciufolini, *Org. Lett.* **17**, 2424–2427 (2015).
21. D. Baidilov *et al*., *Angew. Chem. Int. Ed.* **57**, 10994–10998 (2018).
22. K. Nishikawa *et al*., *Org. Lett.* **23**, 1703–1708 (2021).
23. J. G. Robins, J. S. Johnson, *Org. Lett.* **24**, 559–563 (2022).
24. R. J. Bergeron, J. S. McManis, *J. Org. Chem.* **52**, 1700–1703 (1987).
25. D. B. Konrad, B. Kicin, D. Trauner, *Synlett* **30**, 383–386 (2019).
26. J. Eames *et al*., *J. Chem. Soc., Perkin Trans. 1* **1999**, 1095–1104 (1999).
27. R. G. Soengas, A. M. S. Silva, *Eur. J. Org. Chem.* **2013**, 5022–5027 (2013).
28. T. Mukaiyama, T. Hoshino, *J. Am. Chem. Soc.* **82**, 5339–5342 (1960).
29. A. Hager, N. Vrielink, D. Hager, J. Lefranc, D. Trauner, *Nat. Prod. Rep.* **33**, 491–522 (2016).
30. Y. Basel, A. Hassner, *Synthesis* **1997**, 309–312 (1997).
31. K. N. Houk *et al*., *J. Am. Chem. Soc.* **106**, 3880–3882 (1984).
32. S. Diethelm, E. M. Carreira, *J. Am. Chem. Soc.* **137**, 6084–6096 (2015).
33. C. Cavedon *et al*., *Org. Lett.* **23**, 514–518 (2021).
34. B. M. Trost, Y. H. Rhee, *J. Am. Chem. Soc.* **121**, 11680–11683 (1999).
35. F. E. McDonald, K. S. Reddy, Y. Díaz, *J. Am. Chem. Soc.* **122**, 4304–4309 (2000).
36. M. J. Zacuto, D. Tomita, Z. Pirzada, F. Xu, *Org. Lett.* **12**, 684–687 (2010).
37. S. Beligny, S. Eibauer, S. Maechling, S. Blechert, *Angew. Chem. Int. Ed.* **45**, 1900–1903 (2006).
38. B. Plietker, *J. Org. Chem.* **68**, 7123–7125 (2003).
39. D. W. Knight, I. R. Morgan, A. J. Proctor, *Tetrahedron Lett.* **51**, 638–640 (2010).
40. K. Tsuda *et al*., *Chem. Pharm. Bull.* **12**, 1357–1374 (1964).
41. M. Yotsu-Yamashita, J.-H. Jang, Y. Cho, K. Konoki, *Forensic Toxicol.* **29**, 61–64 (2011).
42. K. Nakanishi, *Bioorg. Med. Chem.* **13**, 4987–5000 (2005).
43. I. J. Curcino Vieira, R. Braz-Filho, *Stud. Nat. Prod. Chem.* **33**, 433–492 (2006).
44. D. B. Konrad *et al*., Supplementary NMR-package for "A Concise Synthesis of Tetrodotoxin," Zenodo (2022). https://doi.org/10.5281/zenodo.6629561.

ACKNOWLEDGMENTS

We thank I. Žamarija, A. Novak, C. Wanzke, K. Schwärzer, E. Miller, R. Bechtel, N. Kurrle, and B. Kicin for experimental assistance and P. Mayer for x-ray analyses. **Funding:** We gratefully acknowledge financial support by the NSF (CHE-1900154). D.B.K. thanks the Friedrich-Ebert-Stiftung for a doctoral scholarship. K.P.R. and B.E.H. are supported by an NYU MacCracken Fellowship. H.A. was supported by the JSPS Overseas Challenge Program for Young Researchers. **Author contributions:** D.T. conceived and directed the project. B.S.M., D.B.K., K.P.R., and D.T. conceptualized and designed the synthetic strategy. B.S.M., D.B.K., K.P.R., H.A., and B.E.H. performed and analyzed the synthetic chemistry experiments. N.S. and K.N.H. performed computational studies. B.S.M., D.B.K., K.P.R., B.E.H., and D.T. prepared the manuscript. B.S.M., D.B.K., and D.T. acquired funding for the project. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Crystallographic data for compounds **11** (CCDC 2117035) and **14** (CCDC 2117036) are available free of charge from the Cambridge Crystallographic Data Centre. Zipped free induction decay files for the NMR spectra of compounds **24**, **25**, **26**, **27**, **S4**, **S5**, and **TTX** have been deposited in Zenodo (*44*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn0571
Materials and Methods
NMR spectra
X-ray data
Calculations
Figs. S1 to S12
Tables S1 to S5
References (*45*–*66*)

Submitted 31 October 2021; accepted 15 June 2022
10.1126/science.abn0571

BIOGEOGRAPHY

Interspecific competition limits bird species' ranges in tropical mountains

Benjamin G. Freeman[1,2]*, Matthew Strimas-Mackey[3], Eliot T. Miller[3]

Species' geographic ranges are limited by climate and species interactions. Climate is the prevailing explanation for why species live only within narrow elevational ranges in megadiverse biodiverse tropical mountains, but competition can also restrict species' elevational ranges. We test contrasting predictions of these hypotheses by conducting a global comparative test of birds' elevational range sizes within 31 montane regions, using more than 4.4 million citizen science records from eBird to define species' elevational ranges in each region. We find strong support that competition, not climate, is the leading driver of narrow elevational ranges. These results highlight the importance of species interactions in shaping species' ranges in tropical mountains, Earth's hottest biodiversity hotspots.

Species that live on tropical mountains usually occur in narrow elevational ranges, whereas species in temperate mountains tend to have broader elevational ranges (*1*, *2*). This pattern is important for determining global patterns of biodiversity because the notable species richness observed in many tropical mountains is due to nearly complete species turnover between low and high elevations (*1*–*3*). It is not known, however, whether elevational ranges in the tropics are more constrained by climate or species interactions.

[1]Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. [2]Department of Zoology, University of British Columbia, Vancouver, BC, Canada. [3]Cornell Lab of Ornithology, Ithaca, New York, USA, 14850.
*Corresponding author. Email: freeman@zoology.ubc.ca

The dominant explanation for narrow elevational ranges in the tropics is that they are the result of physiological adaptation to the low-temperature seasonality of tropical climates. Temperatures range from hot in the lowlands to cold in the highlands but are relatively constant at any single elevation, unlike temperate mountains which experience seasonal temperature fluctuations. Janzen (*4*) was the first to describe how stable temperatures in the tropics could shape species' elevational ranges. He hypothesized that tropical species experience selection to physiologically adapt to the particular thermal conditions they experience, resulting in the evolution of narrow fundamental niches that manifest as restricted elevational ranges (*4*, *5*). In support of this hypothesis, tropical species have thermal tolerances that tend to match the temperature conditions they experience within their elevational zone, and also have narrower thermoneutral zones than their temperate counterparts (*6*–*9*, but see *10*, *11*). Janzen's hypothesis has taken on new relevance in the era of climate change as it predicts that, because tropical species' distributions are limited by physiological adaptation to temperature, they will have disproportionately strong distributional responses to warming temperatures. This geographic prediction is met: Tropical montane species are tracking temperature increases with upslope range shifts much more closely than temperate montane species (*12*).

However, there is an alternative explanation for narrow elevational ranges in the tropics that emphasizes the importance of species interactions rather than climate. At the same time that Janzen put forth his pioneering ideas about temperature seasonality, other researchers were arguing that interspecific competition could limit tropical montane species' elevational ranges (*13*, *14*). In this view, historical and ecological factors explain why large and topographically complex tropical montane regions have accumulated exceptional biodiversity over long time scales (*15*, *16*). The buildup of high regional species richness is hypothesized to result in intense interspecific competition that constrains species to narrow ranges despite their ability to live in a broader array of environments. That is, species have narrow realized niches rather than narrow fundamental niches. The strongest evidence presented for the interspecific competition hypothesis has been case examples of "natural experiments" that compared species' elevational ranges in montane regions where they were sympatric versus allopatric with a closely related species.

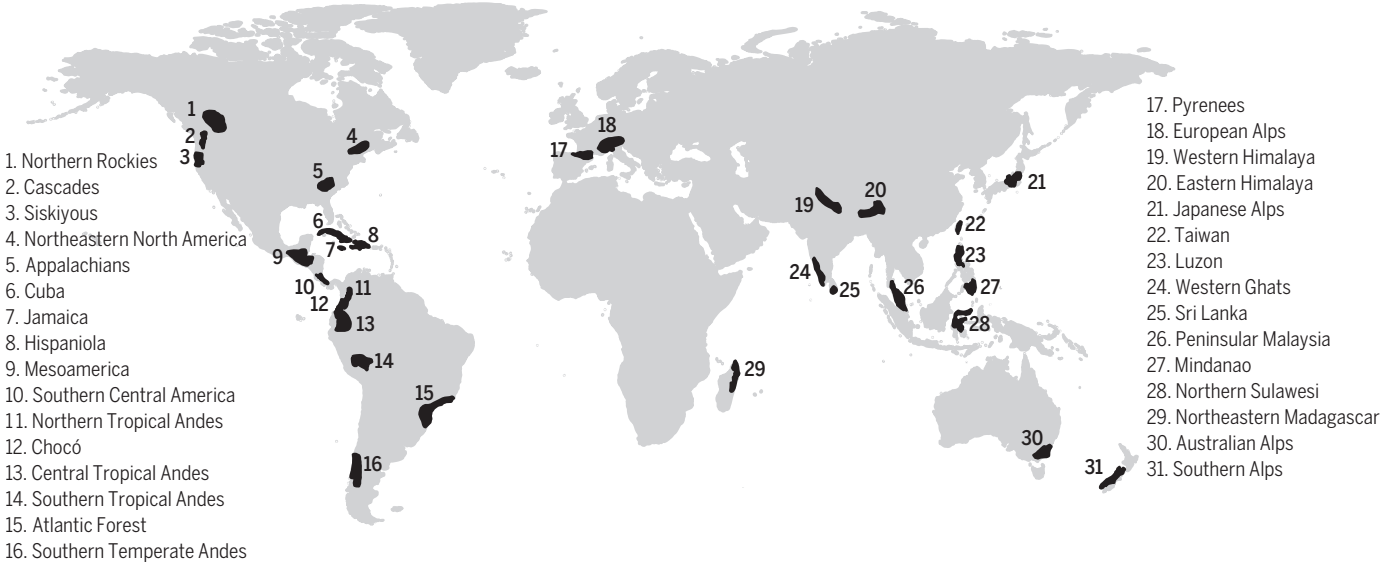


Fig. 1. Map of 31 montane regions included in this study. We included montane regions that had elevational gradients of ≥1400 m, natural forest vegetation along the entire elevational gradient, and sufficient fine-scale distributional data from eBird to define species' regional elevational ranges (mean incidence records per region = 235,438; mean incidence records per 100 m elevational band = 8426).

In many instances such species were reported to be "elevational replacements" in sympatry (i.e., inhabiting narrow nonoverlapping elevational ranges) whereas in allopatry they were reported to live within expanded elevational ranges. These patterns were interpreted as indicating competitive release in allopatry and competitive exclusion in sympatry (*13*, *14*). However, there have been no general tests of the interspecific competition hypothesis.

We provide a global test of contrasting hypotheses to explain why tropical species have narrow elevational ranges: Janzen's hypothesis, which emphasizes abiotic controls on biodiversity, and the interspecific competition hypothesis, which emphasizes biotic controls on biodiversity. We conducted a comparative analysis of forest bird species' elevational ranges within 31 montane regions across the globe (Fig. 1). We defined species' elevational distributions within each region using 4.4 million fine-scale locality records from eBird, a global citizen science project (*17*); all told, our dataset contains elevational ranges for 5397 unique species-by-region combinations (see table S1 for information on regions and fig. S1 for illustration of how we defined species' elevational ranges within regions). Regions ranged in latitude from 43°S to 52°N and in species richness from 23 to 618 species; species richness is defined as the total number of forest bird species within a given region in the eBird dataset. As expected, the regions with the highest species richness were all located in the tropics and had low temperature seasonality [absolute latitude and temperature seasonality were tightly correlated; Spearman's r = 0.93; 95% confidence intervals (CI) = 0.86 to 0.99; $P << 0.001$]. However, tropical regions varied by a factor of 20 in their species richness such that the correlation between temperature seasonality and regional species richness was relatively weak [fig. S2; Spearman's r = −0.46 (95% CI = −0.76 to -0.16, P = 0.0090)]. This allowed us to statistically disentangle temperature seasonality from regional species richness.

We tested contrasting predictions of Janzen's hypothesis and the interspecific competition hypothesis. Janzen's hypothesis predicts that, all else equal, elevational range sizes are narrower in regions with reduced temperature seasonality whereas the interspecific competition hypothesis predicts that, all else equal, elevational range sizes are narrower when regional species richness is high. These two hypotheses are not mutually exclusive. We therefore used the relative explanatory power of one predictor variable versus another to

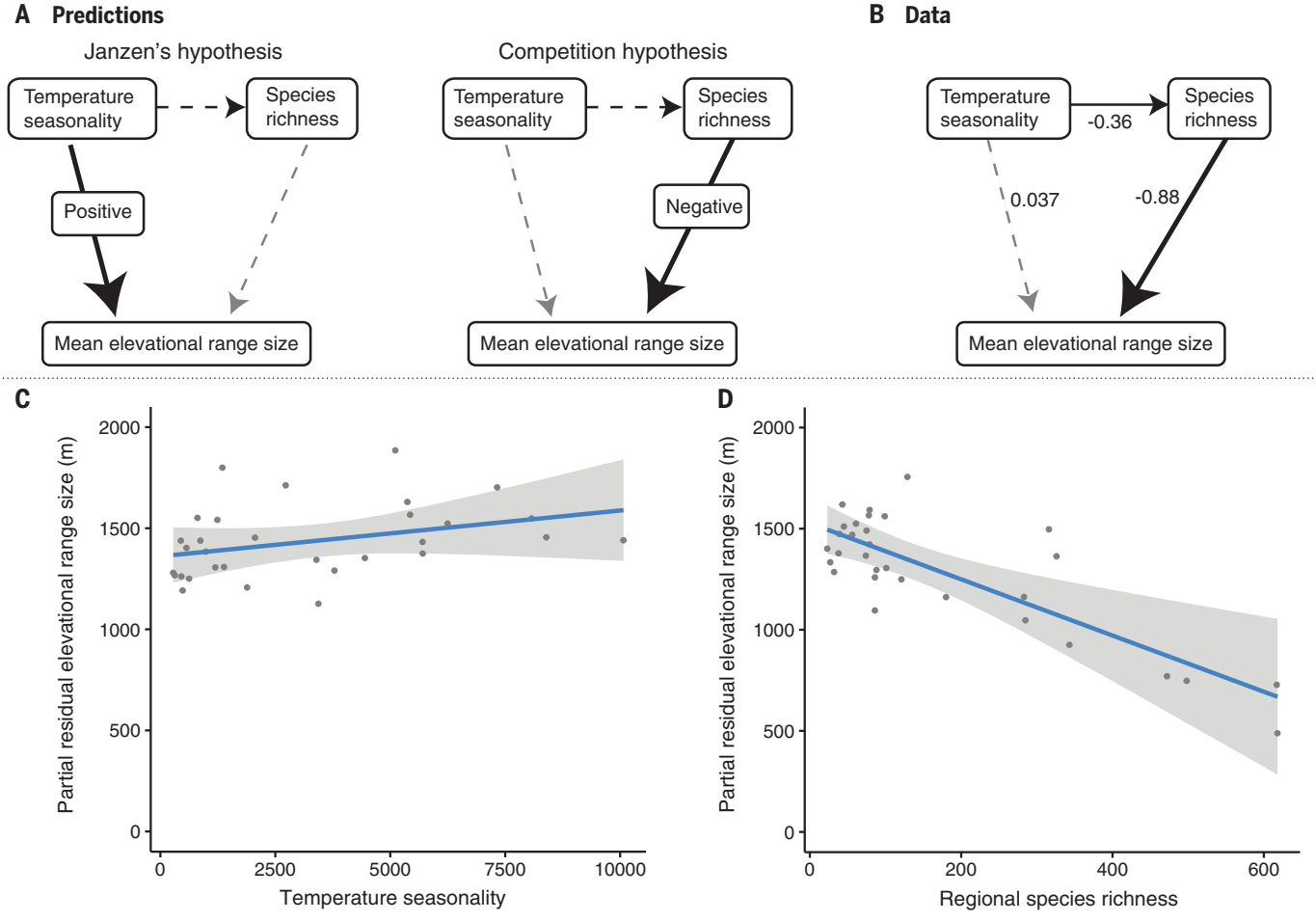


Fig. 2. Predictions and data for the contrasting hypotheses examined in this study. (**A**) Janzen's hypothesis emphasizes climatic controls on species' elevational range sizes whereas the interspecific competition hypothesis emphasizes biotic controls on species' elevational range sizes. We used data from 31 regions to test these non–mutually exclusive predictions by (**B**) path analysis and (**C** and **D**) multiple regression. Elevational range size is better predicted by regional species richness than temperature seasonality in both analyses (path analysis with standardized variables and standard errors: regional species richness = −0.88 ± 0.17, temperature seasonality = 0.037 ± 0.12; multiple regression parameter estimates and standard errors: regional species richness = −1.39 ± 0.37, temperature seasonality = 0.023 ± 0.17). Trendlines in (C) and (D) illustrate expected values with 95% CIs shown in gray shading; points show partial residuals from a multiple regression that included regional species richness, temperature seasonality, and methodological covariates as predictor variables.

assess the relative importance of the two possible mechanisms. All models were multivariate and included mountain height, sampling completeness, and climate change velocity as additional predictor variables.

We found regional species richness to be a better predictor of mean elevational range size of species within a region than temperature seasonality. In a path analysis, regional species richness was negatively associated with elevational range size (−0.88 ± 0.17; parameter estimate and standard error) whereas temperature seasonality was unrelated to elevational range size (0.037 ± 0.12; Fig. 2B, fig. S3, and tables S2 to S4). Similarly, in a multiple regression the evidence for an effect resulting from regional species richness was much stronger (Fig. 2B; regional species richness parameter estimate and standard error = −1.39 ± 0.37, t = −3.74, P = 0.00097; Cohen's f^2 = 0.56) than evidence for an effect of temperature seasonality (Fig. 2C; temperature seasonality parameter estimate and standard error = 0.023 ± 0.017, t = 1.31, P = 0.20; Cohen's f^2 = 0.069; see table S3).

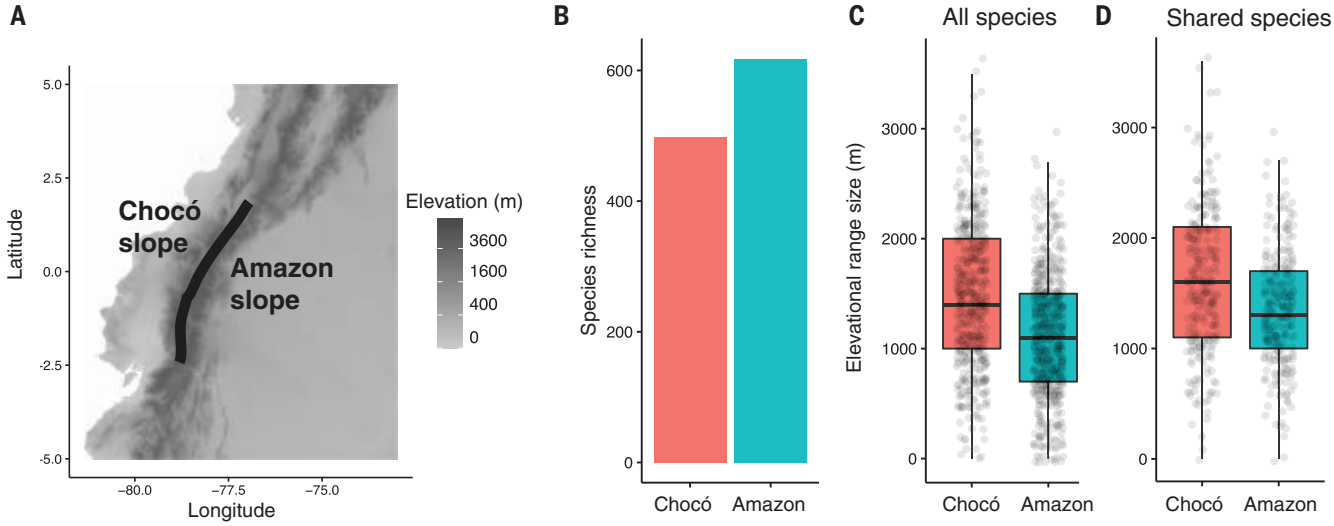


Fig. 3. A case example of how increased regional species richness is associated with narrower elevational ranges. (**A**) At the equator, the tropical Andes are divided into two biogeographic regions: the western Chocó slope and the eastern Amazon slope. (**B**) Species richness of forest birds is high in both regions but is higher on the Amazon slope. Elevational range sizes are larger in the Chocó when considering both (**C**) all species in our analysis (n = 618 species in the Amazon and n = 498 species in the Chocó) and (**D**) the 278 shared species that live in both regions. Effect sizes for (C) and (D) are Cohen's d = 0.61 (95% CI = 0.48 to 0.73) and Cohen's d = 0.43 (95% CI = 0.26 to 0.60), respectively.

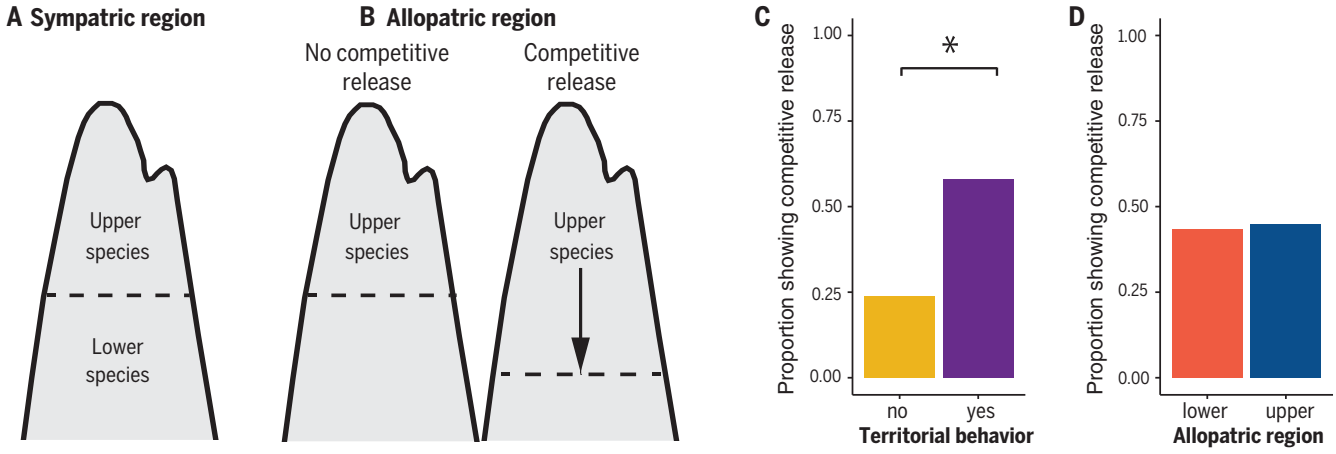


Fig. 4. Pairwise competition between elevational replacements is one mechanism by which higher regional species richness creates stronger interspecific competition that limits species' elevational ranges. (**A**) Elevational replacements are pairs of closely related species that "replace" one another along mountain slopes in sympatric regions. (**B**) We tested for competitive release in allopatric regions, where one species of elevational replacement is not present (A). We asked whether species expanded their distributions in allopatry (with effect sizes Hedge's g > 0.20) to inhabit elevations and positions within multivariate environmental space that in the sympatric region are inhabited by their replacement (sample size = 52 comparisons, see fig. S3 for illustration of competitive release in environmental space; inferences of competitive release in elevation versus environmental space were tightly correlated [table S4]). (**C**) Competitive release was more likely to occur when species defended territories, consistent with the hypothesis that behavioral competition limits elevational ranges in sympatry (P = 0.023). However, competitive release was not more likely when the upper species lived in the allopatric region (**D**), in contrast to the longstanding idea that competition limits warm range limits more so than cold range limits (P = 1; see figs. S7 to S58 for detailed results of each comparison). Effect sizes for (C) and (D) are Cramer's V = 0.34 (95% CI: 0.086 to 0.59) and Cramer's V = 0.013 (95% CI: 0.0014 to 0.31).

Results were similar in an alternative phylogenetic mixed model that analyzed each specific species-by-site combination and incorporated phylogenetic nonindependence between species and spatial nonindependence between regions (table S5). Limited sampling can lead to the systematic underestimation of species' elevational range sizes in lowland and foothill tropical regions where diversity is highest (*18*), but our analysis of sampling completeness finds patterns opposite of that expected if sampling bias affects our results (fig. S4).

The greater importance of regional species richness compared with temperature seasonality holds in a species-level analysis. Species that live within multiple regions in our dataset tended to have smaller elevational ranges in more diverse regions but did not tend to have smaller elevational ranges in regions with lower temperature seasonality. For 176 species found in five or more regions, the average correlation between elevational range size and regional species richness was negative [mean Spearman's r = −0.17 (95% CI = −0.23 to −0.095), degrees of freedom (df) = 175, t = −4.58, P = 0.0000087], but the average correlation between elevational range size and temperature seasonality was close to zero [mean Spearman's r = -0.050 (95% CI −0.13 to 0.027), df = 175, t = −1.29, P = 0.22].

We illustrate these patterns by highlighting two regions on opposite slopes of the Andes at the equator: the western (Chocó) versus the eastern (amazonian) slope (Fig. 3). These adjacent slopes are at the same latitude and have similar climates and elevational relief, but are biogeographically distinct and differ in species richness. The Chocó slope is a center of endemism and highly biodiverse but has fewer species of forest birds than the amazonian slope (in our dataset, 498 versus 618 species of forest-dwelling birds; Fig. 3B). Consistent with predictions of the interspecific competition hypothesis, elevational ranges are narrower on the amazonian slope than the Chocó slope [Fig. 3C; average elevational range sizes = 1114 m versus 1475 m, respectively; df = 971.9, t = 9.91, $P < 2.2^{-16}$; Cohen's d = 0.61 (95% CI = 0.48 to 0.73)]. Again, this pattern is replicated in a species-level analysis. For 278 shared species found in both regions, elevational ranges were more narrow on average on the amazonian slope than on the Chocó slope [Fig. 3D; 1354 m versus 1602 m; df = 277, t = −6.39, P = 7.1^{-10}; Cohen's d = 0.43 (95% CI = 0.26 to 0.60)]. The comparison of the Chocó versus Amazon slopes illustrates our broader result, that regional species richness predicts forest bird species' elevational range sizes within and between regions. We interpret this as indicating that general patterns of elevational specialization in birds are better explained by the interspecific competition hypothesis than by Janzen's hypothesis, while also acknowledging that multiple historical, evolutionary, and ecological factors contribute to the observed patterns.

To shed light on how greater regional species richness generates increased interspecific competition that results in narrow elevational ranges, we examined the mechanism that originally inspired the interspecific competition hypothesis: pairwise competition between "elevational replacements", defined as closely related species that inhabit different elevational zones in sympatry. Elevational replacements are common in a wide range of tropical taxa, but have been best studied in tropical birds (*13*, *14*). We followed previous researchers by analyzing natural experiments, comparing species' ranges in locations where a given species lives in sympatry with a putative competitor with locations where the species lives in allopatry (*19*–*21*). However, we wanted to test for general patterns, which required moving beyond the handful of previously examined case examples that likely suffer from serious ascertainment bias. We therefore conducted a comparative analysis of all 52 natural experiments in Neotropical birds that met predefined criteria, using eBird data to define species' elevational distributions in sympatric and allopatric regions.

We found evidence for competitive release in nearly half of natural experiments: in 23 of 52 cases, species in allopatry inhabited elevational zones and positions within environmental space that in sympatry are occupied by their close relative (elevational release inferred when Hedges' g > 0.20; Fig. 4B, figs. S5 and S6, and table S6). We then tested whether these elevational shifts in allopatry had consequences for elevational range sizes in allopatry. For cases in which we inferred competitive release, elevational ranges expanded in allopatry [fig. S5B; mean elevational range expansion = 330 m (95% CI: 157 to 504 m)], but there was no change in elevational range size between sympatry and allopatry for cases in which we did not infer competitive release [mean elevational range expansion = 2 m (95% CI: −178 to 187 m)]. Thus, elevational replacements often—but certainly not always—have narrow elevational ranges in sympatry that appear to be limited by pairwise competition.

Why do some elevational replacements show competitive release in allopatry whereas others do not? We tested two hypotheses: (i) behavioral interactions are a key mechanism by which competition restricts elevational ranges (*22*–*26*) and (ii) competition restricts species' warm range limits more than their cool range limits (*27*–*30*). Consistent with the hypothesis that behavior is an important mechanism limiting ranges in elevational replacements, competitive release in allopatry is 2.4 times more likely when species defend territories (18 out of 31 cases) than when they do not [5 out of 21 cases; Fig. 4C, Fisher's exact test, P = 0.023; Cramer's V = 0.34 (95% CI: 0.086 to 0.59)]. Field studies measuring interspecific territoriality can test this interpretation. By contrast, competitive release in allopatry was not associated with the relative position of species' elevational ranges [Fig. 4D, Fisher's exact test, P = 1; Cramer's V = 0.013 (95% CI: 0.0014 to 0.31); see figs. S7 to S58 for results of individual natural experiments].

Tropical mountains are home to the greatest concentration of terrestrial biodiversity on Earth because species live only within narrow elevational zones, creating high species turnover along mountain slopes. The prevailing hypothesis for why tropical species live in narrow elevational ranges is that they have evolved physiological adaptations to specific thermal conditions, ultimately leading to the buildup of high species richness in tropical mountains. We present evidence that overturns this explanation—our analysis of a global dataset of millions of citizen science data records reveals that the narrow elevational ranges of tropical birds are driven more by species interactions than by the direct effects of climate. Whether the patterns we demonstrate generalize to other taxa, particularly ectotherms, is a key unanswered question. Regardless, our results suggest a new interpretation for why tropical montane birds (and potentially other tropical taxa) are shifting noticeably upslope (*12*): warming likely causes these upslope shifts indirectly, by altering the outcomes of species interactions, rather than directly through physiological stress. In this view, it is biodiversity itself that makes Earth's hottest biodiversity hotspots disproportionately responsive to climate change.

REFERENCES AND NOTES

1. C. M. McCain, *Ecol. Lett.* **12**, 550–560 (2009).
2. C. D. Cadena *et al.*, *Proc. Biol. Sci.* **279**, 194–201 (2012).
3. I. Quintero, W. Jetz, *Nature* **555**, 246–250 (2018).
4. D. H. Janzen, *Am. Nat.* **101**, 233–249 (1967).
5. C. K. Ghalambor, R. B. Huey, P. R. Martin, J. J. Tewksbury, G. Wang, *Integr. Comp. Biol.* **46**, 5–17 (2006).
6. N. R. Polato *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12471–12476 (2018).
7. J. M. Sunday, A. E. Bates, N. K. Dulvy, *Nat. Clim. Chang.* **2**, 686–690 (2012).
8. A. A. Shah *et al.*, *Funct. Ecol.* **31**, 2118–2127 (2017).
9. B. A. Gill *et al.*, *Proc. Biol. Sci.* **283**, 20160553 (2016).
10. G. A. Londoño, M. A. Chappell, J. E. Jankowski, S. K. Robinson, *Funct. Ecol.* **31**, 204–215 (2017).
11. H. S. Pollock, J. D. Brawn, Z. A. Cheviron, *Funct. Ecol.* **35**, 93–104 (2021).
12. B. G. Freeman, Y. Song, K. J. Feeley, K. Zhu, *Ecol. Lett.* **24**, 1697–1708 (2021).
13. J. M. Diamond, *Science* **179**, 759–769 (1973).
14. J. Terborgh, J. S. Weske, *Ecology* **56**, 562–576 (1975).
15. C. Rahbek *et al.*, *Science* **365**, 1108–1113 (2019).
16. D. Schluter, M. W. Pennell, *Nature* **546**, 48–55 (2017).
17. B. L. Sullivan *et al.*, *Biol. Conserv.* **142**, 2282–2292 (2009).
18. R. K. Colwell, G. C. Hurtt, *Am. Nat.* **144**, 570–595 (1994).
19. R. P. Anderson, M. Gomez-Laverde, A. T. Peterson, *Glob. Ecol. Biogeogr.* **11**, 131–141 (2002).
20. E. E. Gutiérrez, R. A. Boria, R. P. Anderson, *Ecography* **37**, 741–753 (2014).
21. C. D. Cadena, B. A. Loiselle, *Ecography* **30**, 491–504 (2007).
22. B. G. Freeman, *Glob. Ecol. Biogeogr.* **29**, 171–181 (2020).

23. J. E. Jankowski, S. K. Robinson, D. J. Levey, *Ecology* **91**, 1877–1884 (2010).
24. B. Pasch, B. M. Bolker, S. M. Phelps, *Am. Nat.* **182**, E161–E173 (2013).
25. B. G. Freeman, J. A. Tobias, D. Schluter, *Ecography* **42**, 1832–1840 (2019).
26. D. L. Altshuler, *Am. Nat.* **167**, 216–229 (2006).
27. C. Darwin, *On the Origin of Species by Means of Natural Selection* (Murray, 1859).
28. R. H. MacArthur, *Geographical Ecology: Patterns in the Distribution of Species* (Princeton Univ. Press, 1972).
29. A. M. Louthan, D. F. Doak, A. L. Angert, *Trends Ecol. Evol.* **30**, 780–792 (2015).
30. A. Paquette, A. L. Hargreaves, *Ecol. Lett.* **24**, 2427–2438 (2021).

ACKNOWLEDGMENTS

S. Freeman, M. Pennell, L. Rowe, D. Schluter, O. Robinson, R. Colwell, and an anonymous reviewer provided comments that greatly improved the manuscript. We are indebted to eBird users across the globe, as well as the Cornell Lab of Ornithology team that developed and maintains this phenomenal project. **Funding:** This work was funded by the following: Natural Sciences and Engineering Council 379958 and National Science Foundation 1523695 (to B.G.F.). **Author contributions:** Conceptualization: B.G.F., E.T.M., and M.S.-M. Analysis: B.G.F., M.S.-M., and E.T.M. Writing: B.G.F., E.T.M., and M.S.-M. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** Data and code are available at: 10.5281/zenodo.6450245. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl7242
Materials and Methods
Figs. S1 to S58
Tables S1 to S6
References (*31*–*47*)

Submitted 31 July 2021; resubmitted 2 January 2022
Accepted 7 June 2022
10.1126/science.abl7242

CORONAVIRUS

Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution

Tyler N. Starr[1]*†, Allison J. Greaney[1,2,3]†, William W. Hannon[1,4], Andrea N. Loes[1,5], Kevin Hauser[6], Josh R. Dillen[6], Elena Ferri[6], Ariana Ghez Farrell[1], Bernadeta Dadonaite[1], Matthew McCallum[7], Kenneth A. Matreyek[8], Davide Corti[9], David Veesler[5,7], Gyorgy Snell[6], Jesse D. Bloom[1,2,5]*

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has evolved variants with substitutions in the spike receptor-binding domain (RBD) that affect its affinity for angiotensin-converting enzyme 2 (ACE2) receptor and recognition by antibodies. These substitutions could also shape future evolution by modulating the effects of mutations at other sites—a phenomenon called epistasis. To investigate this possibility, we performed deep mutational scans to measure the effects on ACE2 binding of all single–amino acid mutations in the Wuhan-Hu-1, Alpha, Beta, Delta, and Eta variant RBDs. Some substitutions, most prominently Asn^{501}→Tyr (N501Y), cause epistatic shifts in the effects of mutations at other sites. These epistatic shifts shape subsequent evolutionary change—for example, enabling many of the antibody-escape substitutions in the Omicron RBD. These epistatic shifts occur despite high conservation of the overall RBD structure. Our data shed light on RBD sequence-function relationships and facilitate interpretation of ongoing SARS-CoV-2 evolution.

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike receptor-binding domain (RBD) has evolved rapidly since the virus emerged (*1*). We previously used deep mutational scanning to experimentally measure the impact of all single–amino acid mutations on the angiotensin-converting enzyme 2 (ACE2)–binding affinity of the ancestral Wuhan-Hu-1 RBD (*2*). These measurements have helped inform surveillance of SARS-CoV-2 evolution. For example, we identified the N501Y mutation as enhancing ACE2-binding affinity before the emergence of this consequential mutation in the Alpha variant (*3*). (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 mutants, other amino acids were substituted at certain locations; for example, N501Y indicates that asparagine at position 501 was replaced by tyrosine.)

However, as proteins evolve, the impacts of individual amino acid mutations can shift, a phenomenon known as epistasis (*4*). For example, the same N501Y mutation that enhances SARS-CoV-2 binding to ACE2 severely impairs ACE2 binding by SARS-CoV-1 and other divergent sarbecoviruses (*5*). Furthermore, N501Y epistatically enabled other affinity-enhancing mutations that emerged in the Omicron variant of SARS-CoV-2 (*6*–*8*). To more systematically understand how epistasis shifts the effects of mutations, we performed deep mutational scans to measure the impacts of all individual amino acid mutations in SARS-CoV-2 variant RBDs.

We constructed comprehensive site-saturation mutagenesis libraries in the ancestral Wuhan-Hu-1 RBD (201 residues) and RBDs from four variants: Alpha (N501Y), Beta (K417N+E484K+N501Y), Delta (L452R+T478K), and Eta (E484K). We cloned these mutant libraries into a yeast-surface display platform and determined the impact of every amino acid mutation on ACE2 binding affinity and yeast surface-expression levels by means of fluorescence-activated cell sorting (FACS) and high-throughput sequencing (figs. S1 and S2 and data S1) (*2*). The effect of each mutation on ACE2 binding is shown in Fig. 1, and an interactive version of this figure is available at https://jbloomlab.github.io/SARS-CoV-2-RBD_DMS_variants/RBD-heatmaps. We used monomeric ACE2 ectodomain to measure 1:1 binding affinities, which provide more granularity to reveal affinity-enhancing effects compared with our previous measurements using the natively dimeric ACE2 ligand, where some mutational effects are masked by avidity (fig. S1F) (*2*). Mutant effects on ACE2 binding and protein expression in yeast-displayed RBD have been shown to closely correlate with ACE2 binding and protein expression in the context of full spike trimers displayed on mammalian cells (*9*, *10*).

We identified sites where the impacts of mutations differ between RBD variants (Fig. 2 and figs. S3 and S4), reflecting epistasis among the substitutions that distinguish SARS-CoV-2 variants and other mutations across the RBD. These epistatic shifts in mutational effects on ACE2 binding are primarily attributable to the N501Y mutation: The effects of mutations in the Delta (L452R+T478K) and Eta (E484K) RBDs are similar to those in the ancestral Wuhan-Hu-1 RBD, and the differences in the Beta (K417N+E484K+N501Y) RBD largely recapitulate those in the Alpha RBD that contain N501Y alone (Fig. 2, A and B). One exception is a distinct epistatic shift in the effects of mutations to serine or threonine at site 419 in the Beta RBD that introduce an N-linked glycosylation motif when an asparagine is present through the K417N mutation (fig. S3D).

[1]Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. [2]Department of Genome Sciences, University of Washington, Seattle, WA 98109, USA. [3]Medical Scientist Training Program, University of Washington, Seattle, WA 98109, USA. [4]Molecular and Cellular Biology Graduate Program, University of Washington, Seattle, WA 98109, USA. [5]Howard Hughes Medical Institute, Seattle, WA 98109, USA. [6]Vir Biotechnology, San Francisco, CA 94158, USA. [7]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. [8]Department of Pathology, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA. [9]Humabs BioMed SA, a subsidiary of Vir Biotechnology, 6500 Bellinzona, Switzerland.
***Corresponding author. Email: tstarr@fredhutch.org (T.N.S.); jbloom@fredhutch.org (J.D.B.)**
†These authors contributed equally to this work.

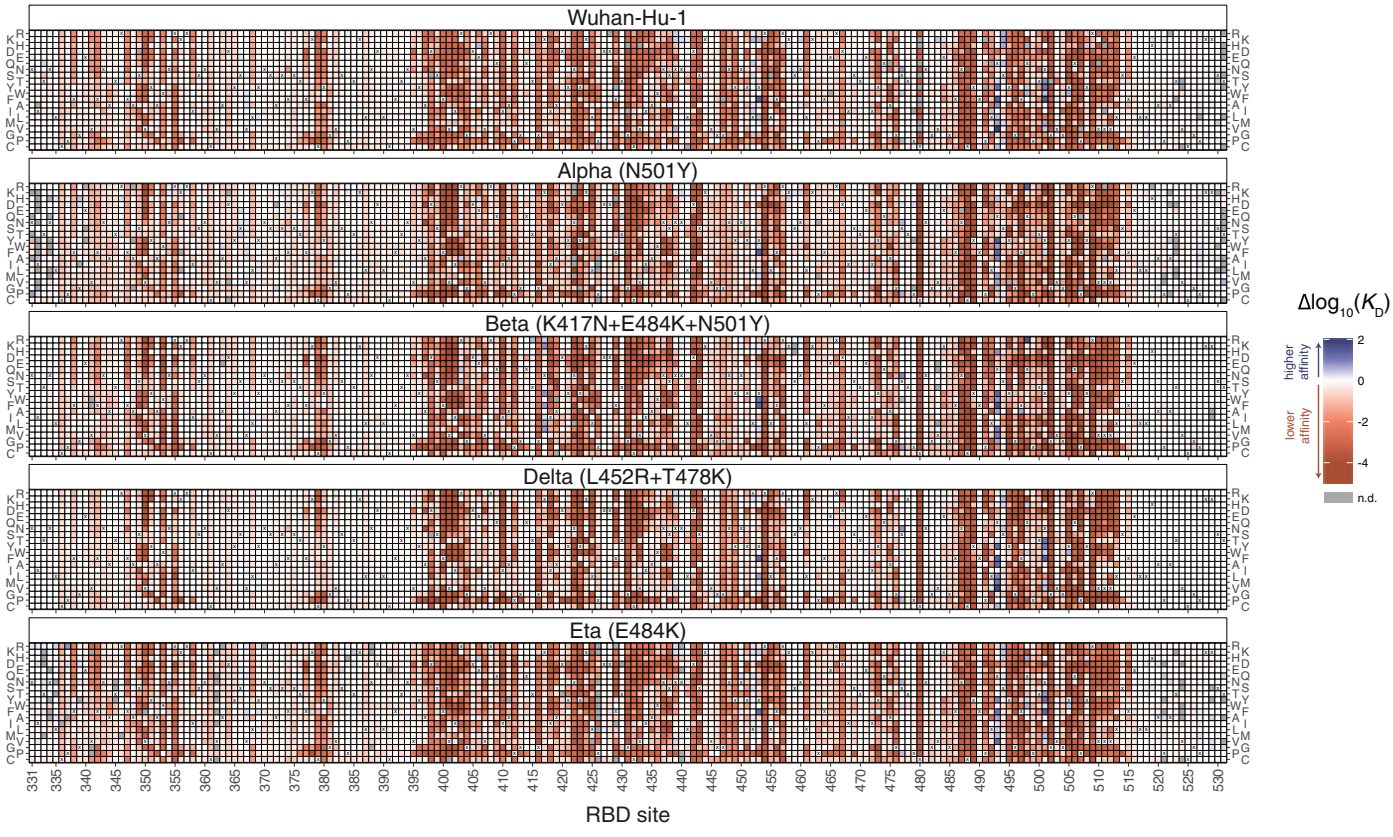


Fig. 1. Deep mutational scanning maps of ACE2-binding affinity for all single–amino acid mutations in five SARS-CoV-2 RBD variants. The impact on ACE2 receptor-binding affinity [$\Delta\log_{10}(K_d)$, where K_d is the dissociation constant] of every single–amino acid mutation in SARS-CoV-2 RBDs, as determined with high-throughput titration assays (fig. S1). The wild-type amino acid in each variant is indicated with an "x", and gray squares indicate missing mutations in each library. An interactive version of this map is at https://jbloomlab.github.io/SARS-CoV-2-RBD_DMS_variants/RBD-heatmaps, and raw data are in data S1. The effects of mutations on RBD surface expression are in fig. S2.

The RBD sites that exhibit notable epistatic shifts because of N501Y fall into three structural groups (Fig. 2B). The largest shift in mutational effects is at the direct N501-contact residue Q498 (Fig. 2C), together with further epistatic shifts at sites 491 to 496 composing the central β strand of the ACE2 contact surface (Fig. 2B and fig. S3A). A second cluster of sites exhibiting epistatic shifts in the presence of N501Y include 446, 447, and 449, which do not directly contact N501 but are spatially adjacent to residue 498 (Fig. 2, B and C, and fig. S3B). A third group of sites that epistatically shift because of N501Y includes residue R403 (Fig. 2C), together with several residues (505, 506, and 406) that structurally link site 501 to site 403 (fig. S3C).

Some of these epistatic shifts are of clear relevance during the evolution of SARS-CoV-2. One of the strongest epistatic shifts is the potentiation of Q498R by N501Y (Figs. 2C and 3A). Although Q498R alone weakly reduces ACE2 affinity in the Wuhan-Hu-1 RBD, it confers a 25-fold enhancement in affinity when present in conjunction with N501Y (which itself improves binding 15-fold in Wuhan-Hu-1), so that the double mutant has a 387-fold increased binding affinity. The Q498R+N501Y double mutation was first discovered in directed evolution studies (*6*) and is present in the RBD of the Omicron BA.1 and BA.2 variants (*8*). The epistasis between these two mutations is crucial for enabling the Omicron RBD to bind ACE2 with high affinity despite having a large number of mutations (*11*–*13*). Specifically, the set of mutations in the Omicron RBD are predicted to strongly impair ACE2 affinity on the basis of their summed single-mutant effects in Wuhan-Hu-1 (Fig. 3B, left), but their summed single-mutant effects in the Beta background (which has N501Y) is about zero (Fig. 3B, right), which is consistent with the actual affinity of the Omicron RBD for ACE2. Therefore, the affinity buffer conferred by the epistatic Q498R+N501Y pair enables the Omicron spike to tolerate other mutations that decrease ACE2 binding (Fig. 3B and fig. S5A) but contribute to antibody escape (fig. S5, B and C) (*14*). Consistent with these affinity measurements, introducing R498Q and Y501N reversions into the Omicron BA.1 spike reduces cell entry by spike-pseudotyped lentiviral particles, suggesting that the remaining Omicron RBD mutations are deleterious without buffering by Q498R+N501Y (Fig. 3C and fig. S6, A and B).

There is also evolutionary relevance of the epistasis of N501Y with mutations on the 446–449 loop, which composes the epitope for an important class of human antibodies (*15*, *16*). Although mutations to G446 escape this class of antibodies in the Wuhan-Hu-1 RBD (*16*, *17*), these mutations incur stronger ACE2-binding deficits in the N501Y background (figs. S3B and S5D). Conversely, mutations to Y449 strongly decrease ACE2-binding affinity in the Wuhan-Hu-1 RBD but are better tolerated when accompanied by N501Y (Figs. 2C and 3D). Mutations to Y449 can escape monoclonal antibodies (fig. S6, C to E) (*15*, *18*) and reduce neutralization by polyclonal sera (*19*, *20*) and have been described in several variants that also contain N501Y, including the C.1.2, A.29, and B.1.640 lineages (*19*, *21*).

To more systematically examine how epistatic shifts caused by N501Y affect patterns of sequence variation during SARS-CoV-2 evolution, we counted the occurrence of substitutions

on a global SARS-CoV-2 phylogeny (*22*). Substitutions more often occurred in backgrounds that contain the amino acid at site 501 with which they had more favorable epistasis with respect to ACE2 affinity (Fig. 3E). Therefore, epistatic shifts caused by N501Y have directly affected patterns of mutation accumulation in prior SARS-CoV-2 evolution, and our data enable identification of mutations such as those at site Y449 whose evolutionary relevance may grow if N501Y variants continue to predominate. Q498R had not previously occurred disproportionately on Y501 genomes until its predominance in Omicron lineages. We hypothesize that the strong affinity gain caused by the Q498R+N501Y double mutant (Fig. 3A) is not directly advantageous itself but rather becomes beneficial in Omicron because

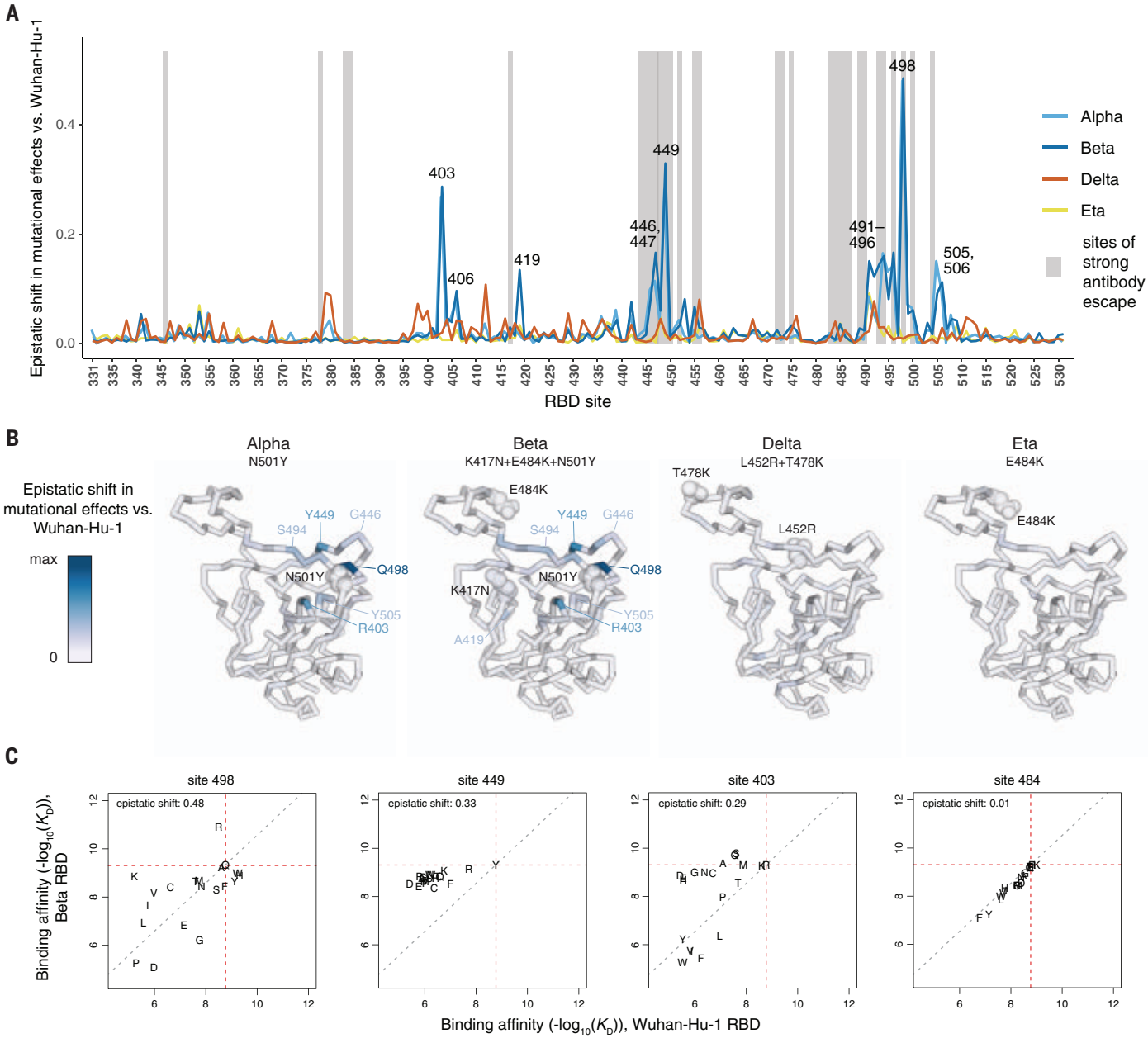


Fig. 2. Epistatic shifts in mutational effects across RBD variants. (**A**) The shift in mutational effects on ACE2 binding at each RBD site between the indicated variant and Wuhan-Hu-1. An interactive version of this plot is at https://jbloomlab.github.io/SARS-CoV-2-RBD_DMS_variants/epistatic-shifts. The epistatic shift is calculated as the Jensen-Shannon divergence in the set of Boltzmann-weighted affinities for all amino acids at each site. Gray shading indicates sites of strong antibody escape based on prior deep mutational scanning of the Wuhan-Hu-1 RBD (*11*). (**B**) Ribbon diagram of the Wuhan-Hu-1 RBD structure (PDB 6M0J) colored according to epistatic shifts. Labeled spheres indicate residues that are mutated in each RBD variant. (**C**) Mutation-level plots of epistatic shifts at sites of interest. Each scatter plot shows the measured affinity of all 20 amino acids in the Beta versus Wuhan-Hu-1 RBD. Red dashed lines indicate the parental RBD affinities, and the gray dashed line indicates the additive (nonepistatic) expectation. Epistatic shifts can reflect idiosyncratic mutation-specific shifts (such as site 498) or global changes in mutational sensitivity at a site (such as site 449). Site 484 does not have a substantial epistatic shift and is shown for comparison. Scatterplots of additional sites of interest are available in fig. S3. Epistatic shifts in mutational effects on RBD expression are available in fig. S4.

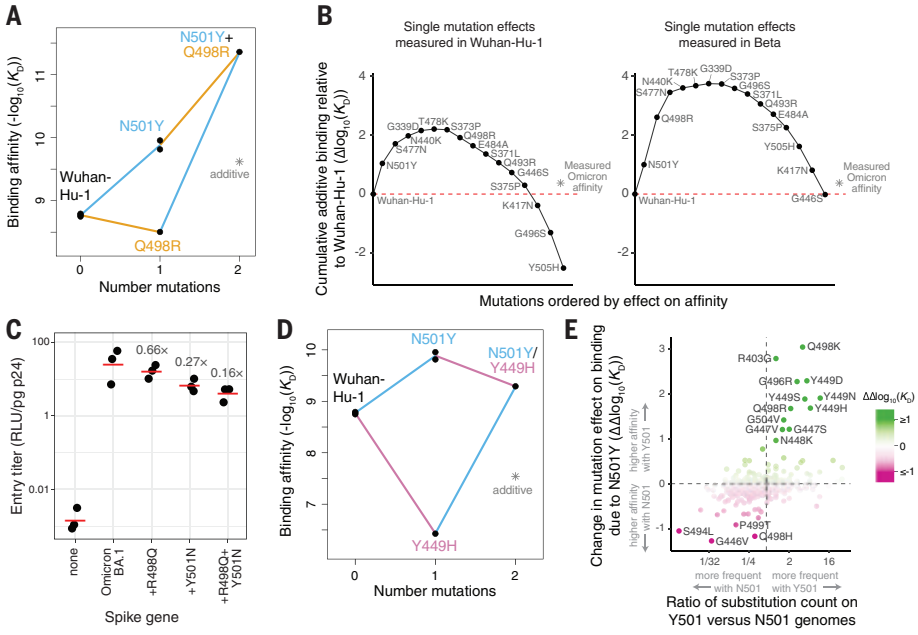


Fig. 3. Functional and evolutionary relevance of epistatic interactions. (**A**) Double-mutant cycle diagram illustrating the positive epistasis interaction between N501Y and Q498R. Asterisk indicates expected double-mutant binding affinity assuming additivity. (**B**) Affinity-buffering of Omicron BA.1 mutations. Each diagram shows the cumulative addition of individually measured effects on ACE2-binding affinity [$\Delta\log_{10}(K_d)$] for each single-RBD substitution in Omicron BA.1 as measured in the Wuhan-Hu-1 (left) or Beta (right) RBDs. Mutation effect is calculated in the labeled direction even when the reference state in a background differs; for example, N501Y in the Beta background is the opposite-sign effect of the measured Y501N mutation. The red line indicates the Wuhan-Hu-1 affinity, and asterisks indicate the actual affinity of the Omicron BA.1 RBD relative to Wuhan-Hu-1 as measured in (*12*) (fig. S5, A to C). (**C**) Efficiency of entry of Omicron BA.1 (or reversion mutant) spike-pseudotyped lentivirus on a human embryonic kidney (HEK)–293T cell line expressing low levels of ACE2 (fig. S6, A and B). Labels indicate fold decrease in geometric mean (red bar) of biological triplicate measurements. (**D**) Double-mutant cycle illustrating positive epistasis between N501Y and Y449H. (**E**) Impact of epistasis on SARS-CoV-2 sequence evolution. Plot illustrates the change in a mutation's effect between Alpha (N501Y) versus Wuhan-Hu-1 deep mutational scanning data, versus the ratio in number of observed occurrences of the substitution in genomes containing N501 versus Y501 in a global SARS-CoV-2 phylogeny as of 25 May 2022 (*22*). We are counting substitution occurrence as an event on the phylogeny independent of the number of offspring of a node and not the raw number of sequenced genomes with which a mutation is observed. A pseudocount was added to all substitution counts to enable ratio comparisons, and substitutions that were observed less than two times in total are excluded. Color scale reinforces the $\Delta\Delta\log_{10}(K_d)$ metric on the *y* axis. Labeled mutations are those with $|\Delta\Delta\log_{10}(K_d)| > 0.9$. The vertical line at $x \sim 0.6$ marks equal relative occurrence on Y501 versus N501 genomes given the larger number of substitutions that had been observed on N501 genomes.

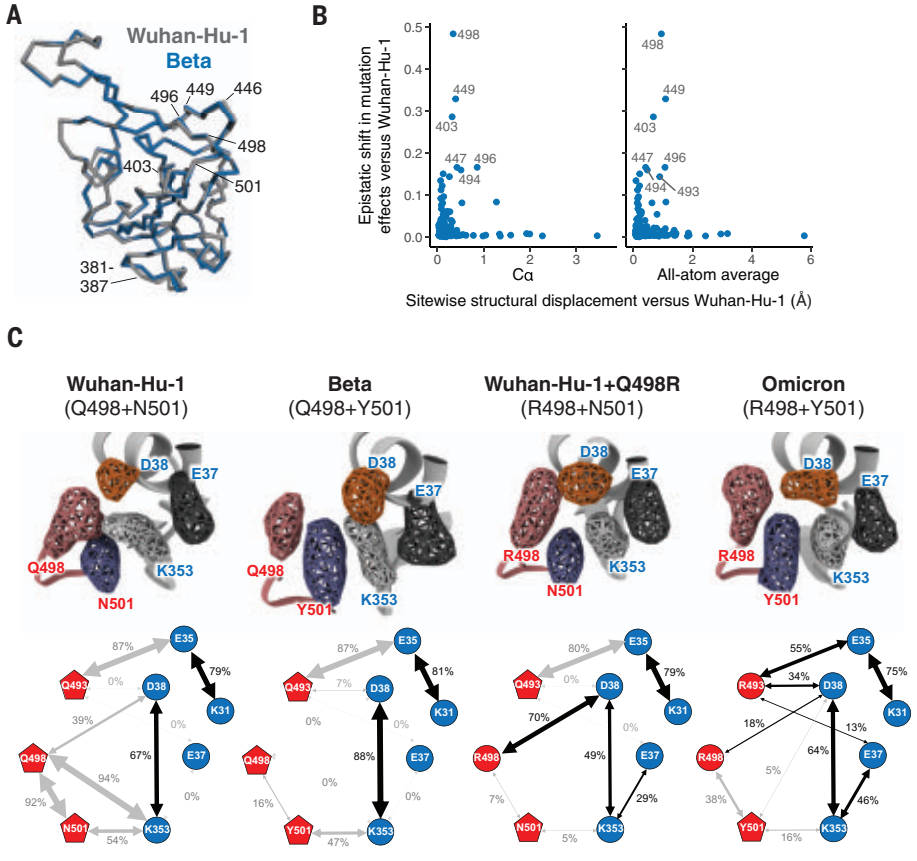


Fig. 4. Epistatic shifts are not accompanied by large structural perturbations. (**A**) Global alignment of the Wuhan-Hu-1 (PDB 6M0J) and Beta (PDB 7EKG) RBD backbones. Key sites are labeled. (**B**) Correlation between the extent of epistatic shift in mutational effects at a site and its structural perturbation in Beta versus Wuhan-Hu-1 RBDs (backbone Cα or all-atom average displacement from aligned x-ray crystal structures) (figs. S7 and S8). (**C**) Molecular dynamics simulation of RBD variants bound to ACE2. Volumetric maps (top) show the 3D space occupied by key residues over the course of simulation. Cartoon diagrams (bottom) illustrate the fraction of simulation frames in which a salt bridge (black arrow) or polar or nonpolar (gray arrow) contact is formed between residue pairs (fig. S9C). Equivalent diagrams for Omicron+Y501N are provided in fig. S9A (R498-N501 for comparison with Wuhan-Hu-1+Q498R), and apo ACE2 is provided in fig. S9B. Histograms of contact distances over the course of the simulations are provided in fig. S9C.

it can buffer other beneficial antibody-escape mutations as described above.

Other common combinations of mutations are not involved in specific epistatic interactions. For example, substitutions at sites 417, 484, and 501 arose together in the Beta and Gamma variants. Early studies disagree on whether there is epistasis among these mutations with respect to ACE2 binding (*6*, *23*, *24*), but our data demonstrate strict additivity (figs. S3E and S5E). The co-occurrence of mutations at these three sites in SARS-CoV-2 variants may instead reflect antigenic selection for E484 and K417 mutants [which escape different classes of neutralizing antibodies (*15*)], whereas N501Y might globally compensate for the affinity-decreasing effect of K417 mutations. These examples illustrate how N501Y can enable viral evolution through specific epistatic modulation (such as Y449 mutations) as well as nonspecific affinity-buffering (such as K417N).

To examine the structural basis for epistatic shifts in mutational effects, we examined ACE2-bound RBD crystal structures of the Wuhan-Hu-1 and Beta RBDs (*25*, *26*), including a newly determined crystal structure of the ACE2-bound Beta RBD (plus antibodies S304 and S309) at 2.45 Å resolution (table S1). These comparisons do not reveal clear structural perturbations that explain epistatic shifts between the Wuhan-Hu-1 and Beta RBDs; residues with large epistatic shifts between backgrounds show extents of variation between Wuhan-Hu-1 and Beta structures similar to those that these residues show within replicate structures of Wuhan-Hu-1 or Beta itself (fig. S7). More broadly, there is minimal change between Wuhan-Hu-1 and Beta RBD backbones (Fig. 4A and fig. S8A), and we did not detect any correlation between structural displacement of backbone or side chain atoms in variant RBD structures and epistatic shifts in mutational effects (Fig. 4B and fig. S8, B to E). These observations indicate that epistatic shifts in mutant effects occur despite conservation of the global static RBD structure.

To explore the cause of epistasis between Q498R and N501Y (Fig. 3A), we performed molecular dynamics simulations of the Wuhan-Hu-1 (Q498-N501), Beta (Q498-Y501), and Omicron (R498-Y501) RBDs bound to ACE2 (*14*, *25*), in addition to in silico mutated complexes of Wuhan-Hu-1+Q498R and Omicron+Y501N (Fig. 4C and fig. S9). The Wuhan-Hu-1 structure features a stable polar contact network between ACE2 residues D38 and K353 and RBD residue Q498. The affinity-enhancing N501Y substitution present in Beta repositions $K353_{ACE2}$ in an orientation that reinforces the $D38_{ACE2}$ salt bridge but disrupts all Q498 contacts. By contrast, the affinity-decreasing Q498R mutation alone improves the coordination between residue 498 and $D38_{ACE2}$ but leaves $K353_{ACE2}$ incompletely satisfied. In Omicron, the Q498R and N501Y combination pose $K353_{ACE2}$ in a stable rotamer that maintains the $D38_{ACE2}$ salt bridge and reanimates the $E37_{ACE2}$ salt bridge present in the apo ACE2 structure (fig. S9B) while adding a new minor salt bridge contact between R498 and $D38_{ACE2}$. This complex epistatic reconfiguration of a polar contact network illustrates how the dynamic basis of RBD:ACE2 interaction leads to dynamic evolutionary variability.

Overall, SARS-CoV-2 has explored a diverse set of mutations during its evolution in humans. Our results show how this ongoing evolution is itself shaping potential future routes of change by shifting the effects of key mutations on receptor-binding affinity. Other human coronaviruses have proven adept at escaping from antibody immunity (*27*) because they can undergo extensive evolutionary remodeling of the amino acid sequence of their receptor-binding domain while retaining high receptor affinity (*28*, *29*). Our work provides large-scale sequence-function maps that help to understand how a similar process may play out for SARS-CoV-2.

REFERENCES AND NOTES

1. K. Tao *et al*., *Nat. Rev. Genet.* **22**, 757–773 (2021).
2. T. N. Starr *et al*., *Cell* **182**, 1295–1310.e20 (2020).
3. Y. Liu *et al*., *Nature* **602**, 294–299 (2022).
4. T. N. Starr, J. W. Thornton, *Protein Sci.* **25**, 1204–1218 (2016).
5. T. N. Starr *et al*., *Nature* **603**, 913–918 (2022).
6. J. Zahradník *et al*., *Nat. Microbiol.* **6**, 1188–1198 (2021).
7. N. Bate *et al*., bioRxiv 473975 [Preprint] (2021); doi: 10.1101/2021.12.23.473975.
8. R. Viana *et al*., *Nature* **603**, 679–686 (2022).
9. K. Javanmardi *et al*., *Mol. Cell* **81**, 5099–5111.e8 (2021).
10. K. K. Chan, T. J. C. Tan, K. K. Narayanan, E. Procko, *Sci. Adv.* **7**, eabf1738 (2021).
11. E. Cameroni *et al*., *Nature* **602**, 664–670 (2022).
12. B. Meng *et al*., *Nature* **603**, 706–714 (2022).
13. M. McCallum *et al*., *Science* **375**, 864–868 (2022).
14. A. J. Greaney, T. N. Starr, J. D. Bloom, bioRxiv 471236 [Preprint] (2021), doi:10.1101/2021.12.04.471236.
15. A. J. Greaney *et al*., *Nat. Commun.* **12**, 4196 (2021).
16. A. J. Greaney *et al*., *PLOS Pathog.* **18**, e1010248 (2022).
17. F. Schmidt *et al*., *Nature* **600**, 512–516 (2021).
18. T. N. Starr *et al*., *Nature* **597**, 97–102 (2021).
19. C. Scheepers *et al*., *Nat. Commun.* **13**, 1976 (2021).
20. T. Tada *et al*., *Cell Rep.* **38**, 110237 (2022).
21. P. Colson *et al*., bioRxiv 21268174 [Preprint] (2021); doi: 10.1101/2021.12.24.21268174.
22. J. McBroome *et al*., *Mol. Biol. Evol.* **38**, 5819–5824 (2021).
23. C. Laffeber, K. de Koning, R. Kanaar, J. H. G. Lebbink, *J. Mol. Biol.* **433**, 167058 (2021).
24. M. Yuan *et al*., *Science* **373**, 818–823 (2021).
25. J. Lan *et al*., *Nature* **581**, 215–220 (2020).
26. P. Han *et al*., *Nat. Commun.* **12**, 6103 (2021).
27. R. Eguia *et al*., *PLOS Pathog.* **17**, e1009453 (2020).
28. A. H. M. Wong *et al*., *Nat. Commun.* **8**, 1735 (2017).
29. Z. Li *et al*., *eLife* **8**, e51230 (2019).
30. T. Starr, A. J. Greaney, W. Hannon, J. Bloom, jbloomlab/SARS-CoV-2-RBD_DMS_variants: published version. Zenodo (2022).

ACKNOWLEDGMENTS

We thank the Genomics and Flow Cytometry core facilities at Fred Hutchinson Cancer Research Center, K. Munson at the University of Washington PacBio Sequencing Services, and the Fred Hutchinson Scientific Computing group supported by ORIP grant S10OD028685. We thank R. Eguia for experimental assistance. We thank S. Weber and the library synthesis team at Twist Bioscience for library construction. We thank P. Hernandez and N. Czudnochowski for support with protein production, J. C. Nix for x-ray data collection, and T. I. Croll for help with structure refinement. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the US Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences under contract DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the NIH NIGMS (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. **Funding:** This project has been funded in part with federal funds from the NIAID-NIH under contracts 75N93021C00015, HHSN272201400006C, and R01AI1417097 (to J.D.B.) and DP1AI158186 and HHSN272201700059C (to D.V.). Funding was also provided by a Pew Biomedical Scholars Award (to D.V.), an Investigators in the Pathogenesis of Infectious Disease Awards from the Burroughs Wellcome Fund (to D.V.), Fast Grants (to D.V.), and the Natural Sciences and Engineering Research Council of Canada (to M.M.). T.N.S. is an HHMI Fellow of the Damon Runyon Cancer Research Foundation. D.V. and J.D.B. are investigators of the Howard Hughes Medical Institute. **Author contributions:** T.N.S., A.J.G., and J.D.B. designed the study. T.N.S. and A.J.G. performed deep mutational scanning experiments. T.N.S. analyzed epistasis in the deep mutational scanning data. W.W.H. created interactive data visualizations. T.N.S. and A.N.L. performed pseudotyped lentiviral entry and neutralization experiments, with reagents and assistance from A.G.F., B.D., K.A.M., and D.C.; T.N.S. and W.W.H. analyzed the SARS-CoV-2 evolutionary data. E.F., J.R.D., M.M., D.V., and G.S. determined the Beta x-ray crystal structure. K.H. and G.S. performed and analyzed molecular dynamics simulation. T.N.S., A.J.G., and J.D.B. wrote the initial draft, and all authors edited the final version. **Competing interests:** J.D.B. has consulted for Moderna on viral evolution and epidemiology and consults for Apriori Bio on deep mutational scanning. T.N.S., A.J.G., A.N.L., and J.D.B. receive a share of intellectual property revenue as inventors on Fred Hutchinson Cancer Research Center–optioned technology and patents related to deep mutational scanning of viral proteins. K.H., E.F., J.R.D., D.C., and G.S. are employees of Vir Biotechnology and may hold shares in Vir Biotechnology. **Data and materials availability:** Site saturation mutagenesis libraries are available from Addgene (catalog 1000000182 to 1000000186). Raw sequencing data are on the NCBI SRA under BioProject PRJNA770094, BioSamples SAMN25941479 and SAMN22208699 (PacBio sequencing), and SAMN25944367 (Illumina barcode sequencing). All code and data at various stages of processing is available together with summaries notebooks detailing the computational pipeline at https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS_variants. Final mutant deep mutational scanning phenotypes are available on GitHub (https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS_variants/blob/main/results/final_variant_scores/final_variant_scores.csv) and data S1, and interactive visualizations of key data are available at https://jbloomlab.github.io/SARS-CoV-2-RBD_DMS_variants. Github files are also archived at Zenodo (*30*). Coordinates for the SARS-CoV-2 Beta RBD coordinated with ACE2, S309, and S304 have been deposited in the Protein Data Bank (PDB) with accession code 8DF5.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo7896
Materials and Methods
Figs. S1 to S9
Table S1
References (*31*–*62*)
MDAR Reproducibility Checklist
Data S1

Submitted 25 February 2022; accepted 23 June 2022
10.1126/science.abo7896

OPTICS

Amplified emission and lasing in photonic time crystals

Mark Lyubarov[1,2,3], Yaakov Lumer[1,2], Alex Dikopoltsev[1,2], Eran Lustig[1,2], Yonatan Sharabi[1,2], Mordechai Segev[1,2,4]*

Photonic time crystals (PTCs), materials with a dielectric permittivity that is modulated periodically in time, offer new concepts in light manipulation. We study theoretically the emission of light from a radiation source placed inside a PTC and find that radiation corresponding to the momentum bandgap is exponentially amplified, whether initiated by a macroscopic source, an atom, or vacuum fluctuations, drawing the amplification energy from the modulation. The radiation linewidth becomes narrower with time, eventually becoming monochromatic in the middle of the bandgap, which enables us to propose the concept of nonresonant tunable PTC laser. Finally, we find that the spontaneous decay rate of an atom embedded in a PTC vanishes at the band edge because of the low density of photonic states.

Photonic time crystals (PTCs) are dielectric media with a refractive index that experiences large, ultrafast periodic variations in time (*1*–*5*). Generally, a wave propagating in a medium undergoing an abrupt change in the refractive index experiences time reflection and time refraction. The time reflection is especially interesting because causality imposes that the wave reflected from the temporal interface propagates backward in space rather than in time (*6*). Periodic modulation of the refractive index makes these time reflections and time refractions interfere giving rise to bands and bandgaps in the momentum (*1*, *3*, *4*). The dispersion relation of PTCs seems analogous to spatial photonic crystals (SPCs), in which the refractive index is periodic in space. However, despite the similarity, there are fundamental differences: SPCs are stationary in time so energy conservation governs most processes, whereas in PTCs, energy is not conserved and causality dictates the dynamics in the system. Conversely, waves propagating in SPCs exchange momentum with the spatial lattice, whereas in spatially homogeneous PTCs, momentum is conserved.

The most important feature of PTCs is the existence of a bandgap in momentum, because the modes associated with this gap have two solutions in which the mode amplitude grows or decays exponentially with time, and both solutions are physical. The exponential growth of the gap modes is nonresonant; it occurs for all wave vectors associated with the momentum gaps, which offers an avenue for amplification of radiation by drawing energy from the modulation. PTCs bear some relation to optical parametric amplifiers, but the latter are resonant phenomena: the frequency of the pump is equal to the sum of the frequencies of the signal and idler and phase matching guarantees conservation of momentum, so only a specific wave is amplified. In contradistinction, PTCs display a significant momentum gap in which every wave is amplified. For a detailed comparison of PTCs and optical parametric amplifiers, see (*7*).

Apart from a momentum band structure, the abrupt temporal modulation of the permittivity also opens up new possibilities such as a frequency conversion (*2*), photon pair creation (*8*–*12*), topological temporal edge states (*5*), antireflection temporal coatings (*13*), extreme energy transformations (*14*), interaction with free electrons (*15*), and amplified localization in temporally disordered media (*16*). Experimentally, time refraction has already been observed in photonics (*17*), whereas time reflection has thus far only been observed with water waves (*18*), acoustic waves (*19*), and elastic waves (*20*). This is because of the highly demanding requirements for observing time reflections: The refractive index change should act as a "wall," analagous to a spatial interface causing Fresnel reflection. For light in the near infrared, the modulation should be at few femtosecond rates with an absolute permittivity change of $\Delta\varepsilon > 0.1$, which is difficult to realize in experimental conditions. However, recent progress with epsilon-near-zero materials (*21*–*24*) brings these ideas close to experimental realization (*25*).

The existence of momentum bands and gaps in a PTC raise fundamental questions about the emission of light by a radiation source embedded in a PTC. An analogous study has led to the discovery of the inhibition of spontaneous emission in the bandgap of SPCs (*26*), which has had major consequences, such as thresholdless lasing (*27*, *28*).

Here, we explore the radiation emitted by a radiation source embedded in a PTC. We formulate the quantum theory describing the emission of light by atoms in an excited state and the classical theory of radiating dipoles embedded in PTCs, and show that radiation is always exponentially amplified when associated with the momentum gap and its linewidth becomes narrower with time. This effect allows us to propose nonresonant tunable PTC lasers which draw their energy from the modulation.

Our model consists of a PTC with a source of radiation inside (Fig. 1A). First, we consider an empty PTC medium (no radiation source) and derive the eigenmodes, and then add an arbitrary radiation source. Starting with Maxwell equations with $\varepsilon = \varepsilon(t)$, $\mu = 1$, we can write the wave equation for the magnetic field as follows:

$$\{\partial_t[\varepsilon(t)\partial_t] + c^2k^2\}\boldsymbol{H}_{\boldsymbol{k}} = 0 \qquad (1)$$

where we use a Fourier transform in space because the system is homogeneous and k is a good quantum number. Physically, this means that the eigenmodes are shaped as plane waves, defined by their wave number k. For each k, this equation has two Floquet eigenmodes:

$$H_k^{1,2}(t) = H_{k0}(t)e^{i\omega_k^{1,2}t} \qquad (2)$$

where $\omega_k^{1,2}$ are Floquet quasifrequencies and $H_{k0}(t)$ is a periodic function in time, constructed from harmonics of the modulation period T. We assume that ε is real (i.e., the medium is lossless), so if $H(t)$ is an eigenmode, i.e., solution of (Eq. 1), then so is $H^*(t)$, which means that $\omega_k^1 = -\omega_k^2 = \omega_k$. Solving for the dispersion relation, we find that the dispersion curve forms a band structure (Fig. 1C). In the bands, the frequency ω_k is real and the two modes are oscillating at the same frequency, whereas in the gaps, ω_k has an imaginary part, with one mode exponentially growing with time and the other exponentially decaying. To explore the response of the PTC to the excitation, we add to Eq. (1) a radiation source associated with a temporally dependent current density $\boldsymbol{j}(\boldsymbol{r},t)$:

$$\{\partial_t[\varepsilon(t)\partial_t] + c^2k^2\}\boldsymbol{H}_{\boldsymbol{k}}(t) = 4\pi ic\boldsymbol{k}\tilde{n}\boldsymbol{j}_{\boldsymbol{k}}(t) \qquad (3)$$

where $\boldsymbol{j}_k(t)$ is a Fourier $\boldsymbol{k}$ component of current $\boldsymbol{j}(\boldsymbol{r},t)$. For a point dipole, we assume $\boldsymbol{j}(\boldsymbol{r},t) = \boldsymbol{d}_0\delta(\boldsymbol{r})e^{i\omega t}\theta(t)$, where $\theta(t)$ is a Heaviside step function denoting that the current is turned on at $t = 0$. Physically, the field $\boldsymbol{H}_k(t)$ is the response of the medium to this current. We can express it in a general form through Green's function as follows:

$$\boldsymbol{H}_{\boldsymbol{k}}(t) = 4i\pi c\int_{-\infty}^{\infty} G_k(t,t')\boldsymbol{k}\times\boldsymbol{j}_{\boldsymbol{k}}(\boldsymbol{t}')dt' \qquad (4)$$

[1]Physics Department, Technion – Israel Institute of Technology, Haifa 32000, Israel. [2]Solid State Institute, Technion – Israel Institute of Technology, Haifa 32000, Israel. [3]Physics and Engineering Department, ITMO University, St. Petersburg 197101, Russia. [4]Department of Electrical and Computer Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel.
*Corresponding author. Email: msegev@technion.ac.il

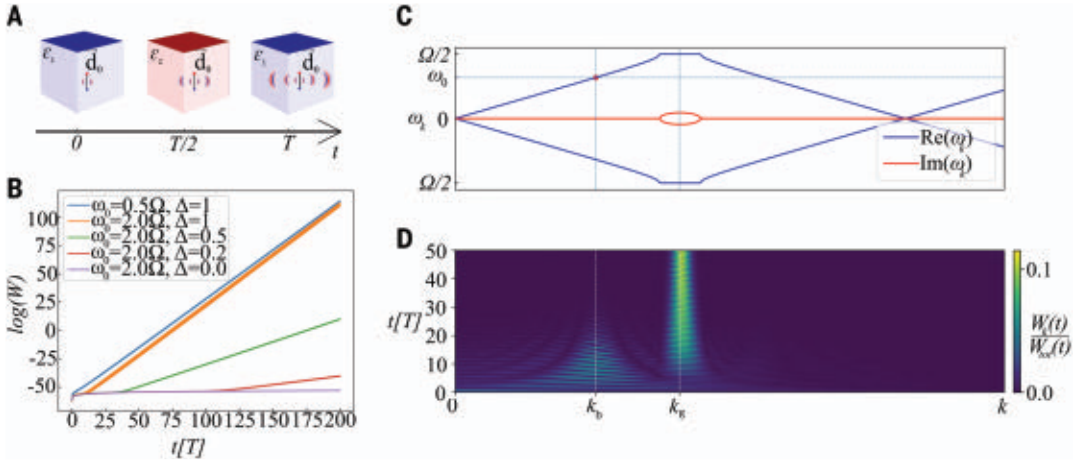


Fig. 1. Emission by a point dipole embedded in a PTC. (**A**) Sketch of the PTC, with permittivity varying as $\varepsilon(t) = \varepsilon_{av} + \Delta/2 \cdot \cos(\Omega t)$, $\Omega = 2\pi/T$ with a dipole antenna inside. The dipole radiation is exponentially amplified with time. (**B**) Exponential growth of electromagnetic energy associated with the dipole emission for different dipole frequencies ω_0 and modulation amplitudes Δ. (**C**) Complex dispersion relation (band structure) of the PTC for $\varepsilon_{av} = 2$, $\Delta = 1$. The values of w_k at the bandgap around k_g are complex, indicating exponentially growing and decaying eigenmodes. (**D**) Power spectrum of dipole emission versus wave number as it evolves with time. Initially, the point dipole with frequency ω_0 excites all eigenmodes. k_0 is a wave number of the mode resonantly excited by the dipole with frequency ω_0: $\omega_k(k_0) = \omega_0$. The emission linewidth initially occupies all bands and gaps, located at k_g, but eventually, after a short time, the radiation in the gap becomes dominant and narrower with time, reflecting the stronger emission at midgap. In each moment in time, the spectrum is normalized by the total radiation power. The horizontal axes in (C) and (D) coincide.

where

$$\left(\partial_t(\varepsilon(t)\partial_t) + c^2k^2\right)G_k(t,t') = \delta(t-t') \quad (5)$$

and then express this Green's function through the eigenmodes from Eq. (2):

$$G_k(t,t') = \begin{cases} 0, & t < t' \\ \dfrac{H^2{}_k(t')H^1{}_k(t) - H^1{}_k(t')H^2{}_k(t)}{\varepsilon(t')(H^2{}_k(t')\partial_{t'}H^1{}_k(t') - H^1{}_k(t')\partial_{t'}H^2{}_k(t'))}, & t > t' \end{cases} \quad (6)$$

Green's function, $G_k(t,t')$, represents the response of the medium at time t to a single homogeneous "flash" at time t'. The detailed derivation of Eq. (6) is provided in (*7*). A closer look at Eqs. 4 to 6 reveals that, in the momentum bandgap where $Im(\omega_k) \neq 0$, the medium responds with exponentially growing emission even to the slightest flash of radiation emitted from the current source. This seemingly counterintuitive feature is a consequence of the lack of energy conservation in the medium. In fact, the energy deposited into the exponentially growing gap modes comes not from the source but rather from the external modulation of the medium. The exponentially growing dipole emission is shown in Fig. 1B for various dipole frequencies and permittivity profiles. The growth rate barely depends on the frequency of the dipole but strongly depends on the amplitude of the permittivity modulation. The larger the modulation, the sooner the growth takes place and the steeper it is. The energy spectrum (k) of the dipole emission and its evolution with time are depicted in Fig. 1D. The numerical simulation of the fields in Fig. 1, B and D, is described in detail in section 4 of (*7*). Initially, the point dipole with frequency ω_0 excites all the eigenmodes with proper wave number k_0 such that $\omega_k(k_0) = \omega_0$. This is because these waves lie on the dispersion curve and thus are perfectly phase matched. However, within a few oscillation cycles, the gap modes start to dominate even if k_0 does not belong to the gap. These modes are not phase matched with the dipole frequency, but they nevertheless grow exponentially in time, which overshadows any phase matching.

We can understand the exponentially growing response in a PTC through Fig. 2, which shows the difference between excited gap modes in SPC and in PTC, where the excitation in the SPC is by a point source in real space and the excitation in the PTC is by a flash in time. The solution of Eq. (5) should be expressed through two eigenmodes on either side of the excitation point and stitched with two stitching conditions. The physical constraints in both cases reveal which contributions are unphysical and should be removed. In the case of the SPC (Fig. 2A), the solution must obey energy conservation, so only evanescent waves are allowed on either side of the excitation point in space. Therefore, the response to the excitation at a frequency in the gap of a SPC are evanescent waves. Conversely, in the PTC (Fig. 2B), two of the four modes are propagating back in time and therefore cannot be excited because they are restricted by causality. Thus, Green's function must be expressed with two forward-propagating waves in time, one of which is exponentially decaying and the other exponentially growing, which is allowed because there is no energy conservation in PTCs.

This analysis explains the exponentially growing dipole emission in a PTC. The dipole excites the gap modes, which, once excited, grow exponentially regardless of the dipole, even when mismatched. The key issue here is that a point dipole excites modes with all k, $\boldsymbol{j_k} \neq 0$ $\forall \boldsymbol{k}$, including the exponentially growing gap modes. Thus, any point source in a PTC results in a exponentially growing emission, even when the excitation is a single flash in time. The emission from this flash will grow exponentially, drawing energy from the modulation.

Next, we quantize our model. First, we write the electromagnetic field Hamiltonian in a PTC

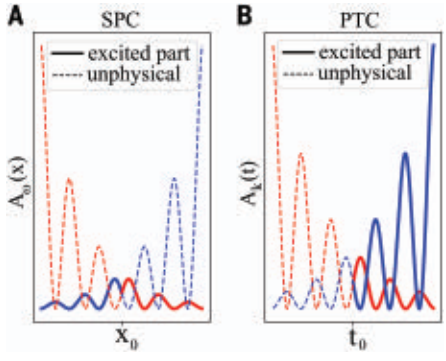


Fig. 2. Excitation of gap modes in SPC and PTC. (**A**) The one-dimensional SPC is excited by a point source at position x_0 and emits at a given frequency within the photonic bandgap. The source can couple only to the spatially evanescent part on either side of x_0 because of energy conservation. (**B**) In the PTC, the source is a flash at t_0 and can excite only the parts of the modes that evolve forward in time, as dictated by causality. One of these two modes is exponentially growing in time.

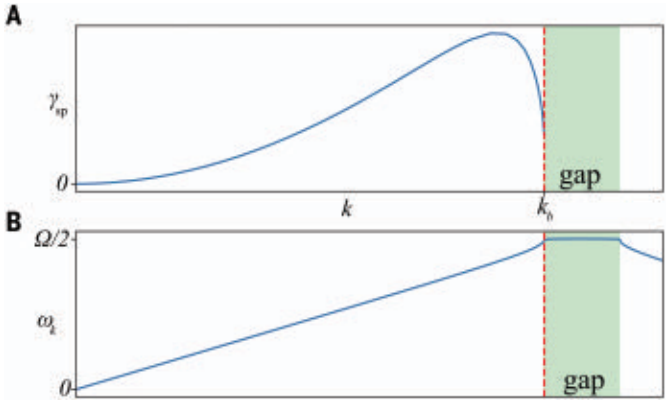


Fig. 3. Spontaneous emission rate into the Floquet modes associated with the first band. (**A**) Spontaneous emission rate in the first band of the PTC. Starting at a low wave number, the emission rate increases and then reaches a maximum, but then declines and goes to zero at the band edge k_b, where the band structure is curved. (**B**) Dispersion in the first band of the PTC. The slope of the bandstructure becomes more vertical closer to the bandgap, indicating the vanishing density of states.

as follows:

$$H_f = \hbar\sum_{\mathbf{k}}\frac{ck}{n(t)}\left(\frac{\frac{n(t)}{n_r}+\frac{n_r}{n(t)}}{2}\left(a^{\dagger}_{\boldsymbol{k}}a_{\boldsymbol{k}}+a^{\dagger}_{-\boldsymbol{k}}a_{-\boldsymbol{k}}\right)+\frac{\frac{n(t)}{n_r}-\frac{n_r}{n(t)}}{2}\left(a_{\boldsymbol{k}}a_{-\boldsymbol{k}}+a^{\dagger}_{\boldsymbol{k}}a^{\dagger}_{-\boldsymbol{k}}\right)\right) \quad (7)$$

where $n(t)=\sqrt{\varepsilon(t)}$ is the time-varying refractive index, n_r is the mean value of refractive index obtained by averaging through one modulation cycle, and $a^{\dagger}_{\boldsymbol{k}}(a_{\boldsymbol{k}})$ are the creation (annihilation) operators for mode with the wave vector $\boldsymbol{k}$. This Hamiltonian is derived in (*7*) following the quantization procedure described in (*29*). It follows our intuition gained in the classical case: It is time dependent through $n(t)$ and it conserves momentum, $\left[H_f,\ \sum_k \boldsymbol{k}a^{\dagger}_k a_k\right]=0$, but it does not conserve the number of photons. The Hamiltonian (Eq. 7) allows describing the dynamics of the free field for each photon pair $\{k, -k\}$ separately. The resulting dynamics agrees with the classical case: For modes with k associated with the band of the PTC, the expectation value of the number of photons, $N_k(t)=\langle\psi(t)|a^{\dagger}_k a_k+a^{\dagger}_{-k}a_{-k}|\psi(t)\rangle$, oscillates near some constant value, whereas if k belongs to the PTC bandgap, N_k grows exponentially with time at the same rate as in the classical case. The periodic variation of $n(t)$ allows introducing the Floquet eigenmodes $|\psi_k(t)\rangle=e^{-i\omega_k t}|\varphi_k(t)\rangle$ of the Hamiltonian (Eq. 7), with ω_k being the Floquet eigenfrequency. Let us first list the main features of the quantum Floquet eigenmodes, the detailed analysis of which is provided in (*7*). In the bands, ω_k coincides with the Floquet frequency calculated in the classical analysis, and the Floquet eigenstates experience weak oscillations in the number of photons. Conversely, in the bandgap, the eigenstates of the Hamiltonian cannot exist: By correspondence with the classical case, N_k in the gap eigenmodes should grow exponentially, which is impossible with Hermitian Hamiltonians such as the one in Eq. (7). The absence of eigenstates in the gap brings complexity in studying the dynamics of the excited atom, interacting with the radiation field described below, but the exponential growth of the number of photons in the momentum gap and the classical/semiclassical intuition allow us to make some safe statements on the dynamics in this unusual quantum system.

To describe the emission from excited atoms in PTC, we add the atomic and the interaction parts to the Hamiltonian (Eq. 7):

$$H = H_f + H_a + H_{\text{int}} \quad (8)$$

$$H_a = \hbar\omega_0\sigma_z \quad (9)$$

$$H_{\text{int}} = \sum_k \frac{\hbar g_k}{\varepsilon(t)}\left(a_k+a^{\dagger}_k\right)\left(\sigma_+ + \sigma_-\right) \quad (10)$$

where we assume a two-level atom and dipole interaction. We first analyze what happens with an initially excited atom interacting with the vacuum field. In the case of a static medium, the result is the exponential decay of the atom from the excited state to the ground state, known as spontaneous emission. In a PTC, no analytic solution is feasible, because the number of photons in the initially empty gap modes grows exponentially regardless of the atom. This means that the atom emission into these modes cannot be clearly divided into spontaneous and stimulated emissions: The rate of transitions grows with time as a consequence of the photons already created by the PTC. In addition to stimulated emission, stimulated absorption also takes place, which results in complex dynamics of the atom. As in the classical case, the growth in the number of photons barely depends on the frequency of atomic transition. Moreover, even if the two-level atom is not in resonance with the momentum-gap, i.e., $\omega_0 \neq \Omega/2$, the emission into gap modes eventually governs the dynamics of the atom.

It is now natural to ask if there are any circumstances under which we can still talk about spontaneous emission (in the usual sense of being induced by quantum fluctuations with no photons around) in a PTC and what the physical consequences of this might be. This question can be answered partially by addressing the Floquet modes associated with the band, ignoring the influence of the gap modes. This assumption can be justified if the decay time of the atom is shorter than the inverse growth rate of the number of photons in the gap modes $\tau_{sp} < 1/Im(\omega_k)$ or if the PTC with the embedded atom is placed in a resonator with all resonator eigenmodes residing inside the PTC bands (rather than in the gaps). In this case, we show in (*7*) that the spontaneous emission rate is:

$$\gamma = \frac{V}{\hbar^2\pi}\sum_m \left|V^m_{fi}\right|^2 \frac{k_m^2}{\left|\frac{\partial\omega_f}{\partial k}\right|_{k=k_m}} \quad (11)$$

where

$$V^m_{fi} = \frac{1}{T}\int_0^T \langle\varphi_f(t)|H_{\text{int}}(t)|\varphi_i(t)\rangle e^{im\Omega t}dt \quad (12)$$

is the coupling constant between the initial and the final Floquet eigenstates through H_{int} and $k_m : \omega_f(k_m)=\omega_0+m\Omega$ is the wave number of the mode corresponding to m^{th} harmonic of the atomic transition (*30*). Analyzing the dynamics of the emission rate $\gamma(k)$ within the band, we observe that there are two competing contributions: the closer to the band edge the larger the $\left|V^m_{fi}\right|$, because for modes in the vicinity of the gap the oscillations are larger, whereas at the band edge, the density of states, $\rho\propto k^2\left(\frac{\partial\omega}{\partial k}\right)^{-1}$ is smaller. Fig. 3A shows that at the band edge, the rate of spontaneous emission vanishes because the density of states goes down to zero. The low density of states is apparent from the vertical slope of the dispersion near the band edge (Fig. 3B). The implication is intriguing: Even though the Floquet modes have larger oscillations closer to the band edge, which naturally increases the strength of the light-matter interaction, the emission rate at the edge goes to zero because there are no states to radiate into. Thus, an "atom" or a nano-antenna with directional emission at the band edge would stay in the excited state forever, unable to relax to the ground state through spontaneous emission.

In one-dimensional PTCs, the presence of a gap in the momentum alters the light-matter interactions in a profound way, bringing to question foundational issues such as the meaning of spontaneous and induced emission in such media and the lifetime of an atom in excited

states. The exponential growth of energy in the modes associated with the PTC gap and the nonmonotonous growth rate raise the exciting idea of PTC lasers extracting their energy from the modulation. The simplest setting for such a laser is to construct a resonator by placing mirrors on either side of the dielectric medium with its permittivity modulated in time. The cavity length should be much larger than the wavelength of the waves of interest, such that momentum conservation applies despite the finite size of the resonator. Cavities with shorter lengths can also exhibit momentum gaps but require additional treatment of the spatial modes. Because the amplification of the waves associated with the gap modes attains a maximum at midgap, any saturation mechanism will eventually result in stable monochromatic emission. Thus, controllable periodic change of the permittivity can give rise to coherent radiation from an almost arbitrary source and, under some conditions, the emission can be shaped into pulses by designing the modulation.

REFERENCES AND NOTES

1. F. Biancalana, A. Amann, A. V. Uskov, E. P. O'Reilly, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **75**, 046607 (2007).
2. J. R. Zurita-Sánchez, P. Halevi, J. C. Cervantes-Gonzalez, *Phys. Rev. A* **79**, 053821 (2009).
3. J. R. Reyes-Ayona, P. Halevi, *Appl. Phys. Lett.* **107**, 074101 (2015).
4. A. M. Shaltout, J. Fang, A. V. Kildishev, V. M. Shalaev, "Photonic time-crystals and momentum band-gaps," paper presented at the 2016 Conference on Lasers and Electro-Optics (CLEO): Science and Innovations, San Jose, CA, 5–10 June 2016.
5. E. Lustig, Y. Sharabi, M. Segev, *Optica* **5**, 1390–1395 (2018).
6. J. T. Mendonça, P. K. Shukla, *Phys. Scr.* **65**, 160–163 (2002).
7. See the supplementary materials.
8. A. B. Shvartsburg, *Phys. Uspekhi* **48**, 797–823 (2005).
9. M. Uhlmann, G. Plunien, R. Schützhold, G. Soff, *Phys. Rev. Lett.* **93**, 193601 (2004).
10. J. T. Mendonca, G. Brodin, M. Marklund, *Phys. Lett. A* **372**, 5621–5624 (2008).
11. J. Sloan, N. Rivera, J. D. Joannopoulos, M. Soljačić, *Phys. Rev. Lett.* **127**, 053603 (2021).
12. F. Belgiorno *et al*., *Phys. Rev. Lett.* **105**, 203901 (2010).
13. V. Pacheco-Peña, N. Engheta, *Optica* **7**, 323–331 (2020).
14. H. Li, S. Yin, E. Galiffi, A. Alù, *Phys. Rev. Lett.* **127**, 153903 (2021).
15. A. Dikopoltsev *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2119705119 (2022).
16. Y. Sharabi, E. Lustig, M. Segev, *Phys. Rev. Lett.* **126**, 163902 (2021).
17. G. Lerosey *et al*., *Phys. Rev. Lett.* **92**, 193904 (2004).
18. V. Bacot, M. Labousse, A. Eddi, M. Fink, E. Fort, *Nat. Phys.* **12**, 972–977 (2016).
19. A. Derode, P. Roux, M. Fink, *Phys. Rev. Lett.* **75**, 4206–4209 (1995).
20. C. Draeger, M. Fink, *Phys. Rev. Lett.* **79**, 407–410 (1997).
21. N. Kinsey *et al*., *Optica* **2**, 616–622 (2015).
22. L. Caspani *et al*., *Phys. Rev. Lett.* **116**, 233901 (2016).
23. M. Z. Alam, I. De Leon, R. W. Boyd, *Science* **352**, 795–797 (2016).
24. O. Reshef, I. De Leon, M. Z. Alam, R. W. Boyd, *Nat. Rev. Mater.* **4**, 535–551 (2019).
25. E. Lustig *et al*., "Towards photonic time-crystals: Observation of a femtosecond time-boundary in the refractive index," paper presented at the 2021 Conference on Lasers and Electro-Optics (CLEO): QELS–Fundamental Science, San Jose, CA, 9–14 May 2021.
26. E. Yablonovitch, *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
27. O. Painter *et al*., *Science* **284**, 1819–1821 (1999).
28. S. Noda, A. Chutinan, M. Imada, *Nature* **407**, 608–610 (2000).
29. W. Vogel, D.-G. Welsch, *Quantum Optics* (Wiley, 2006).
30. T. Bilitewski, N. R. Cooper, *Phys. Rev. A* **91**, 033601 (2015).

ACKNOWLEDGMENTS

Funding: The initial stages of this research were funded by a grant from the US Air Force Office of Scientific Research. **Author contributions:** All authors contributed substantially to all relevant aspects of this research. **Competing interests:** The authors declare no competing interests. **Data availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo3324
Materials and Methods
Figs. S1 to S3
References (*31*–*35*)

Submitted 28 January 2022; accepted 27 May 2022
10.1126/science.abo3324

CORONAVIRUS

Pathogenicity, transmissibility, and fitness of SARS-CoV-2 Omicron in Syrian hamsters

Shuofeng Yuan[1]†, Zi-Wei Ye[1]†, Ronghui Liang[1]†, Kaiming Tang[1]†, Anna Jinxia Zhang[1]†, Gang Lu[2,3]†, Chon Phin Ong[4]†, Vincent Kwok-Man Poon[1,5], Chris Chung-Sing Chan[1,5], Bobo Wing-Yee Mok[1], Zhenzhi Qin[1], Yubin Xie[1], Allen Wing-Ho Chu[1], Wan-Mui Chan[1], Jonathan Daniel Ip[1], Haoran Sun[6], Jessica Oi-Ling Tsang[1,5], Terrence Tsz-Tai Yuen[1], Kenn Ka-Heng Chik[1,5], Chris Chun-Yiu Chan[1], Jian-Piao Cai[1], Cuiting Luo[1], Lu Lu[1,5], Cyril Chik-Yan Yip[6], Hin Chu[1,5,7], Kelvin Kai-Wang To[1,5,6,7,8], Honglin Chen[1,5,6,8], Dong-Yan Jin[4,8]‡, Kwok-Yung Yuen[1,3,5,6,7,8]‡, Jasper Fuk-Woo Chan[1,3,5,6,7,8]‡§*

The in vivo pathogenicity, transmissibility, and fitness of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Omicron (B.1.1.529) variant are not well understood. We compared these virological attributes of this new variant of concern (VOC) with those of the Delta (B.1.617.2) variant in a Syrian hamster model of COVID-19. Omicron-infected hamsters lost significantly less body weight and exhibited reduced clinical scores, respiratory tract viral burdens, cytokine and chemokine dysregulation, and lung damage than Delta-infected hamsters. Both variants were highly transmissible through contact transmission. In noncontact transmission studies Omicron demonstrated similar or higher transmissibility than Delta. Delta outcompeted Omicron without selection pressure, but this scenario changed once immune selection pressure with neutralizing antibodies—active against Delta but poorly active against Omicron—was introduced. Next-generation vaccines and antivirals effective against this new VOC are therefore urgently needed.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first identified in late 2019 and quickly developed into the most important global health challenge in recent decades (*1*–*3*). Despite rapid generation of vaccines and antivirals and global implementation of various nonpharmaceutical public health measures, the coronavirus disease 2019 (COVID-19) pandemic continues nearly 2 years after the emergence of more variants of concern (VOC) exhibiting enhanced immunoevasiveness and/or transmissibility (*4*). The Alpha (B.1.1.7) variant emerged in mid-2020 and quickly outcompeted the Beta (B.1.351) variant (*5*, *6*). The Delta (B.1.617.2) variant with its enhanced transmissibility and moderate level of antibody resistance has subsequently replaced the Alpha variant since mid-2021.

The Omicron (B.1.1.529) variant, first identified in South Africa in November 2021, has now affected at least 149 countries (*7*, *8*). This new VOC has a notably high number of mutations (>30) at the spike, which significantly reduces the neutralizing activity of vaccine-induced serum antibodies as well as therapeutic monoclonal antibodies (*9*–*15*). Preliminary analyses of the severity of infections caused by Omicron compared with previous variants as determined by hospitalization rates have been inconclusive, with some showing reduced

[1]State Key Laboratory of Emerging Infectious Diseases, Carol Yu Centre for Infection, Department of Microbiology, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong Special Administrative Region, China. [2]Key Laboratory of Tropical Translational Medicine of Ministry of Education, Hainan Medical University, Haikou, Hainan, China. [3]Academician Workstation of Hainan Province, Hainan Medical University-The University of Hong Kong Joint Laboratory of Tropical Infectious Diseases, Hainan Medical University, Haikou, Hainan, China; and The University of Hong Kong, Pokfulam, Hong Kong Special Administrative Region, China. [4]School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong Special Administrative Region, China. [5]Centre for Virology, Vaccinology and Therapeutics, Hong Kong Science and Technology Park, Hong Kong Special Administrative Region, China. [6]Department of Infectious Disease and Microbiology, The University of Hong Kong-Shenzhen Hospital, Shenzhen, Guangdong Province, China. [7]Department of Microbiology, Queen Mary Hospital, Pokfulam, Hong Kong Special Administrative Region, China. [8]Guangzhou Laboratory, Guangdong Province, China.
***Corresponding author. Email: jfwchan@hku.hk**
†These authors contributed equally to this work. ‡These authors contributed equally to this work. §Present address: State Key Laboratory of Emerging Infectious Diseases, Carol Yu Centre for Infection, Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong Special Administrative Region, China; and Department of Infectious Disease and Microbiology, The University of Hong Kong-Shenzhen Hospital, Shenzhen, Guangdong Province, China.

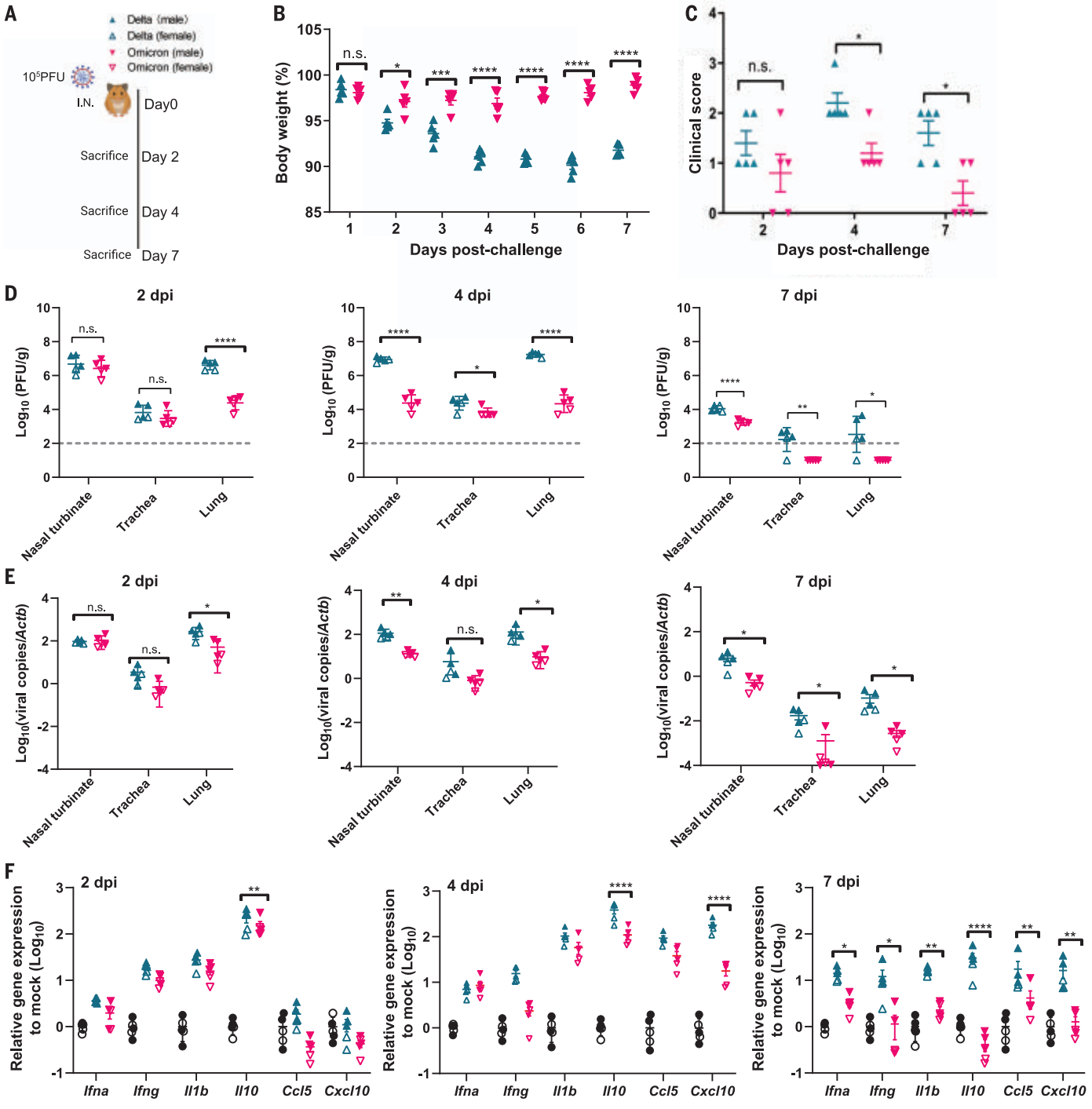


Fig. 1. Pathogenicity of Omicron and Delta in the Syrian hamster model of COVID-19. (**A**) Scheme of the pathogenicity study comparing infections caused by Omicron and Delta in Syrian hamsters. At 0 dpi each hamster was intranasally inoculated with SARS-CoV-2 (n = 15 for each variant). In the first independent experiment, the hamsters (n = 5 for each variant) were kept alive for body weight and clinical score monitoring in (**B** and **C**). (B) Body weight changes (n = 5 male hamsters per variant, Student's t test). (C) Clinical scores. A score of 1 was given for each of the following clinical signs: lethargy, ruffled fur, hunchback posture, and rapid breathing (n = 5 male hamsters per variant, two-tailed Mann-Whitney U test). In the second independent experiment, the hamsters were sacrificed at 2, 4, and 7 dpi for analysis in (**D** to **F**) (n = 5 including 3 male and 2 female hamsters per variant per time point). (D) Respiratory tract tissue infectious virus titers and (E) viral loads (Student's t test). The dotted line in (D) represents the limit of detection of the plaque assay (100 PFU/g). (F) Lung cytokine and chemokine gene expression profiles (Student's t test). Values on the y axis represent the changes in Omicron- or Delta-infected relative to mock-infected samples (drawn in $\log_{10}$ scale). Black dots indicate the mock-infected group (n = 5 per time point, including 3 male hamsters as indicated by black dots and 2 female hamsters as indicated by white dots). Data represent mean ± standard deviation. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001. n.s., not significant. *Actb*, beta-actin; I.N., intranasal.

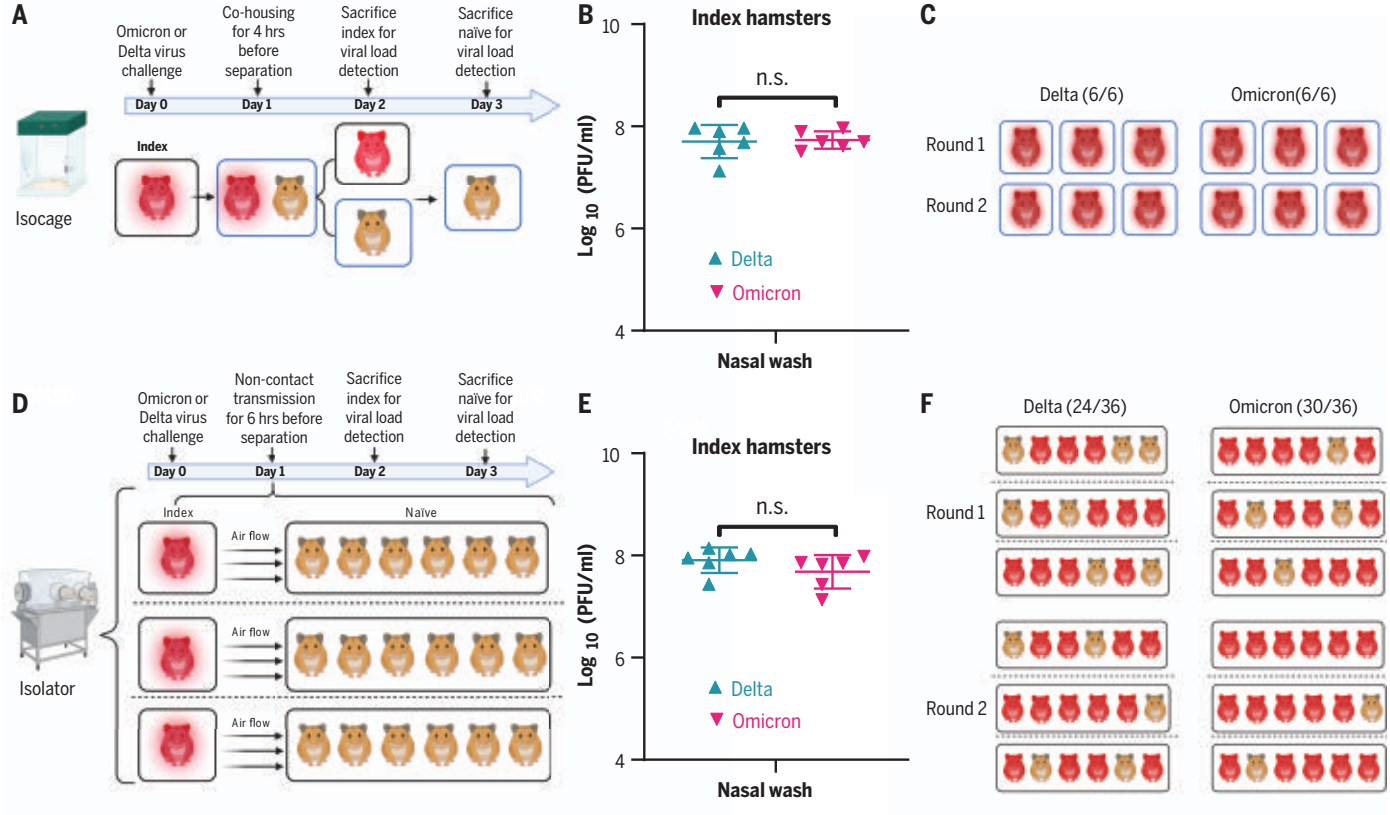


Fig. 2. Contact and noncontact transmission of Omicron and Delta among Syrian hamsters. (**A**) Scheme of the contact transmission study. (**B**) The nasal wash infectious virus titers in the intranasally SARS-CoV-2–challenged index hamsters at 1 dpi were determined by plaque assay to ensure successful infection of animals. Data represent mean ± standard deviation of the pooled results of two independent experiments (n = 3 animals per group per experiment, Student's t test). (**C**) Positive rates of infection among the naïve hamsters after exposure to either Omicron or Delta in two independent experiments (n = 3 animals per group per experiment). Red hamsters indicate those that were infected. (**D**) Scheme of the noncontact transmission study. (**E**) Nasal wash infectious virus titers in the intranasally SARS-CoV-2–challenged index hamsters at 1 dpi. Data represent mean ± standard deviation of the pooled results of two independent experiments (n = 3 animals per group per experiment, Student's t test). (**F**) Positive rates of infection among the naïve hamsters after exposure to either Omicron or Delta (n = 18 animals per group per experiment, Student's t test). Brown hamsters indicate those that were not infected and red hamsters indicate those that were infected, as in (C). In both contact and noncontact transmission studies, hamsters with Ct value ≤40 in either nasal turbinates or lungs were considered infected. n.s., not significant.

hospitalization rates and others showing a lack of significant difference (*16*). What is more apparent from early epidemiological data is that Omicron is spreading rapidly even in populations with high uptake rates of two-dose COVID-19 vaccinations (*17*, *18*). However, whether this is a result of the intrinsic transmissibility of Omicron or other extrinsic environmental and social factors is unknown. At present, the in vivo pathogenicity, transmissibility, and fitness of Omicron is poorly understood. We investigated these virological attributes of Omicron by comparing them with those of Delta in a Syrian hamster model of COVID-19, which closely simulates nonlethal human disease and has been widely used to study various aspects of SARS-CoV-2 infection biology (*19*–*25*).

We first compared the clinical signs, viral burden, and cytokine and chemokine profiles of Omicron and Delta in our hamsters (Fig. 1A), finding that Omicron-infected animals showed limited reductions in body weight (<5%) (Fig. 1B). Furthermore, their clinical scores were significantly lower than those of the Delta-infected hamsters (Fig. 1C). Early after infection [2 days post infection (dpi)], the viral loads and infectious virus titers of the two variants in the nasal turbinates and trachea were similar but the lung viral loads and virus titers were significantly lower in the Omicron- than Delta-infected hamsters (Fig. 1, D and E). During the acute (4 dpi) and regenerative (7 dpi) phases of infection, the viral burden of Omicron became consistently lower than that of Delta throughout the upper and lower respiratory tract (Fig. 1, D and E). At 7 dpi, the viral titers in the trachea and lungs were already below the detection limit [<100 plaque-forming units (PFU)/g] in the Omicron-infected hamsters. Earlier clearance of virus shedding in oral swabs (fig. S1A) and feces (fig. S1B) were observed in the Omicron-infected hamsters. Consistent with these findings, the Omicron-infected hamsters generally expressed lower levels of inflammatory cytokine and chemokine genes (Fig. 1F) and/or proteins (fig. S2) between 2 and 7 dpi. At 7 dpi, the dysregulated inflammatory cytokine and chemokine response was almost completely normalized in the Omicron-infected hamsters. The antibody response against the variant-specific spike receptor-binding domain (RBD) of the Omicron-infected hamsters was also significantly lower than that of the Delta-infected hamsters (fig. S3).

The lung sections of the Omicron-infected hamsters collected at 2 dpi showed alveolar wall congestion whereas Delta-infected hamsters exhibited more severe and diffuse peribronchiolar and alveolar inflammatory infiltrates (fig. S4). At 4 dpi, both groups of hamsters exhibited bronchiolar epithelial destruction and peribronchiolar and perivascular inflammatory infiltrates. However, the pathological changes in the Delta-infected hamsters were more diffuse than the Omicron-infected hamsters. At 7 dpi, the lung sections of the Omicron-infected hamsters appeared mostly normal whereas those of the Delta-infected hamsters

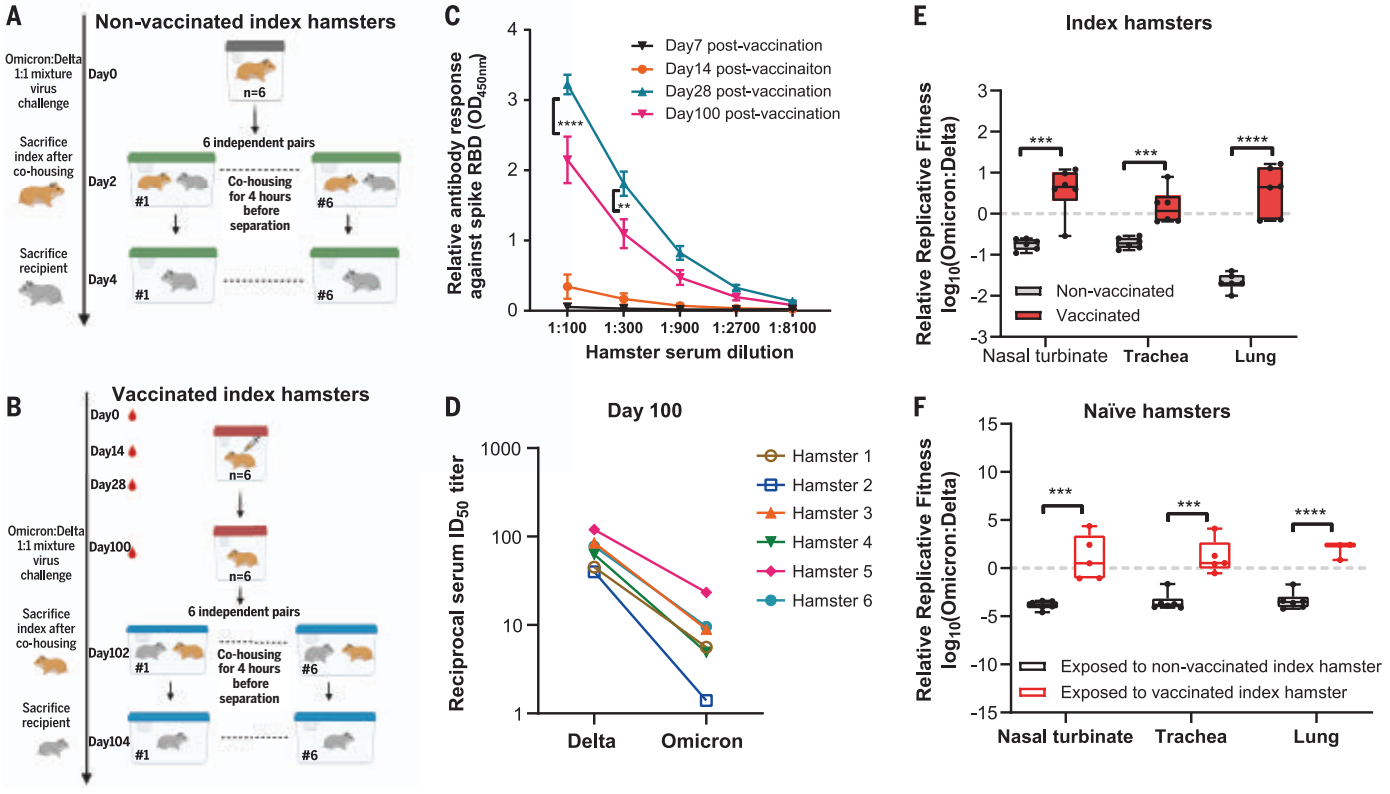


Fig. 3. Comparative in vivo fitness of Omicron and Delta in Syrian hamsters. Schemes of the in vivo competition models with (**A**) nonvaccinated and (**B**) vaccinated index hamsters. (**C**) The hamster serum samples (n = 6) were collected at the indicated days following vaccination for detection of antibody against wild-type SARS-CoV-2 spike receptor-binding domain (RBD) (HKU-001a strain, GenBank accession number: MT230904) with an enzyme-linked immunosorbent assay (Student's t test). (**D**) The neutralizing activity of the vaccinated hamster serum samples (n = 6) collected at day 100 after vaccination against authentic Omicron and Delta using a microneutralization assay. ID_{50}, 50% inhibitory dose. (**E**) The Omicron-to-Delta ratios in the nasal turbinate, trachea, and lung of the nonvaccinated and vaccinated index hamsters, and those of the (**F**) naïve hamsters exposed to the nonvaccinated or vaccinated index hamsters (Student's t test). Data represent mean ± standard deviation of the results of n = 6 biological replicates. $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$.

continued to exhibit blood vessel congestion and alveolar wall inflammatory infiltration. This suggested that lung damage was resolved more quickly in the Omicron-infected hamsters. The histological scores were significantly lower in the Omicron-infected hamsters between 2 and 7 dpi. Additionally, viral nucleocapsid proteins were more abundantly expressed in the lung sections of the Delta-infected than Omicron-infected hamsters throughout 2 to 7 dpi (fig. S5). Similarly, the Omicron-infected hamsters generally showed less severe histopathological changes (fig. S6) and less abundant viral nucleocapsid protein expression (fig. S7) in their nasal turbinates than the Delta-infected hamsters from 4 to 7 dpi. Thus, Omicron exhibits attenuated pathogenicity in Syrian hamsters compared with Delta.

Another key question we sought to answer was the comparative transmissibility of Omicron and Delta in vivo. To this end, we first cohoused six index SARS-CoV-2–challenged hamsters (n = 3 for each variant) with six naïve hamsters for 4 hours in a 1:1 ratio (Fig. 2A). The experiment was repeated twice. All index hamsters had similar nasal wash virus titers at 1 dpi (Fig. 2B). All 12 naïve hamsters were found to be infected 2 days after exposure (Fig. 2C), indicating that both variants are highly transmissible through close contact. The mean virus titer in the lungs of Delta-infected naïve hamsters was significantly higher than that of Omicron-infected naïve hamsters (fig. S8). Next, we randomly grouped 42 hamsters into six groups of index and naïve hamsters (1:6 ratio) in our established noncontact transmission system; we then repeated the experiment twice (total n = 84) (Fig. 2D) (*20*). The hamsters were sacrificed at 2 dpi (index) or 2 days after exposure to index (naïve). All index hamsters were successfully infected with similarly high nasal wash virus titers at 1 dpi (Fig. 2E). All 12 cages, including 30 out of 36 (83.3%) Omicron-exposed and 24 out of 36 (66.7%) Delta-exposed naïve hamsters became infected (Fig. 2F). Although the sample size was underpowered to reach statistical significance (P = 0.173, chi-square test), the transmission rate of Omicron was consistently ~10 to 20% higher than that of Delta in both rounds of experiments if the individual numbers of RT-PCR–positive naïve hamsters were counted.

To investigate why Omicron emerged as the dominant circulating SARS-CoV-2 variant, we compared its fitness with that of Delta. Consistent with our recent preliminary findings at an early time point, Delta consistently exhibited a significant fitness advantage over Omicron for up to 72 hours post infection in vitro (fig. S9, A and B) (*26*). However, this scenario changed when selection pressure by vaccinated sera containing antibodies with reduced anti-Omicron but preserved anti-Delta neutralizing activity was present (fig. S9C), with Omicron significantly outcompeting Delta (fig. S9, D to F). We next validated our in vitro findings with in vivo competition models. We included both nonvaccinated (Fig. 3A) and vaccinated (Fig. 3B) index hamsters and intranasally challenged them with the two variants (1:1 ratio). The vaccinated index hamsters were observed 100 days after vaccination with an inactivated SARS-CoV-2 vaccine and showed waning serum antibody

responses in comparison with the peak activity at 28 days postvaccination (Fig. 3C). Their serum-neutralizing antibody activity against Omicron was markedly lower than Delta (Fig. 3D). In the nonvaccinated index hamsters, Delta significantly outcompeted Omicron. By contrast, Omicron exhibited a marked fitness advantage over Delta in the vaccinated index hamsters (Fig. 3E and fig. S10). Delta similarly outcompeted Omicron in naïve hamsters that were exposed to nonvaccinated index hamsters. By contrast, when naïve hamsters were exposed to vaccinated index hamsters the replication advantage of Delta was significantly diminished (Fig. 3F). Thus, Delta shows a fitness advantage over Omicron in the absence of selection pressure. Under immune selection pressure, however, Omicron becomes the dominant variant causing infection.

Novel SARS-CoV-2 variants will continue to emerge as long as the virus maintains its wide circulation among humans and nonhuman mammals. Although it has recently become clear that Omicron exhibits immune evasion to most existing anti–SARS-CoV-2 therapeutic monoclonal antibodies and vaccine-induced neutralizing antibodies (*9*–*15*), understanding of the in vivo pathogenicity, transmissibility, and fitness of this VOC remains incomplete. In the pathogenicity study we demonstrated that although the viral load and infectious virus titer of the two variants were similar in the nasal turbinates and trachea, Omicron is significantly less replicative in the lungs even at the early post infection stage (2 dpi). Omicron also consistently induces less cytokine and chemokine dysregulation and tissue damage in the lungs. In human lung–derived Calu-3 cells, Omicron shows reduced replication compared with Delta and the D614G strain in pseudovirus and/or live virus assays (*26*, *27*). Moreover, the Omicron spike exhibits reduced receptor binding and fusogenicity as well as S1 subunit shedding in vitro (*26*, *27*). A recent study using an ex vivo lung organ culture model showed that Omicron exhibits enhanced replication compared with Delta in the bronchi (*28*). This differs from the findings of the present study and recently reported animal model data which show that Omicron is generally less replicative than Delta throughout the upper and lower respiratory tract (*29*, *30*). This apparent discrepancy may be caused by different study models and conditions. Our in vivo findings help explain the observations in early epidemiological studies that report lower rates of hospitalization caused by Omicron compared with Delta (*16*, *31*). The shorter duration of virus shedding in oral swabs and feces of Omicron-infected hamsters may also have implications for infection control of Omicron-infected patients should the same viral shedding pattern be confirmed in humans.

The transmissibility of Omicron is a key factor in optimizing public health control measures and predicting the evolution of the pandemic. Recent epidemiological studies have suggested that Omicron may be spreading even faster (up to 4.2 times) than Delta in its early stage (*32*). The estimated effective reproductive number (R_t) of Omicron in South Africa and the UK is 2.5 to 3.7, with a doubling time of 3 days (*33*). Our head-to-head comparison showed that Omicron exhibits similar or higher transmissibility than Delta through both contact and noncontact transmission despite generally having lower respiratory tract viral loads. Other factors such as the efficiency of the variants in entering cells and their ability to remain as infectious particles in aerosols or on inanimate surfaces for prolonged periods should be investigated (*34*).

To provide insight on the replacement of Delta by Omicron as the dominant SARS-CoV-2 variant, we compared their fitness in cell culture and hamster models. Notably, the rapidly disseminating Omicron was consistently outcompeted by Delta in vitro and in nonvaccinated hamsters, which may be the result of its unusually high number of genetic mutations. Omicron exhibits a significant fitness advantage over Delta under selection pressure in vitro in the presence of vaccinated serum and in hamsters with waning serum-neutralizing antibody levels. These findings help explain why Omicron has outcompeted Delta and has become the predominant SARS-CoV-2 strain especially in populations with high rates of previous infection and/or vaccination with first generation COVID-19 vaccines eliciting suboptimal anti-Omicron neutralizing antibody responses. However, our findings should be interpreted carefully and should not be considered as evidence against COVID-19 vaccination. By contrast, our findings in hamsters with waning serum-neutralizing antibody >3 months after vaccination are supportive of booster vaccines because recent data have shown that antibody neutralization is mostly restored by mRNA vaccine booster doses (*35*).

Our study has certain limitations. The transmission rate of SARS-CoV-2 may vary according to different durations of exposure and diagnostic criteria applied. We selected 6 hours of noncontact transmission to simulate the real-life scenarios of RT-PCR testing after being exposed to an infected index patient within the same facility for a routine business day and on medium-haul flights. It would be worthwhile to compare the transmissibility of Omicron and Delta after different durations of exposure; it may also be important to investigate the pathogenicity and transmissibility of Omicron in additional animal models as each model has its own advantages and disadvantages in recapitulating human disease.

In summary, the present study shows that despite comparatively lower pathogenicity than Delta, Omicron undoubtedly still causes obvious disease in infected hosts. Taking into consideration Omicron's high transmissibility, our findings highlight the urgent need for next-generation COVID-19 vaccines and broad-spectrum therapeutics, as well as improve upon nonpharmaceutical measures to reduce acute and chronic disease burden (Long Covid) on the general public and healthcare facilities.

REFERENCES AND NOTES

1. J. Cohen, D. Normile, *Science* **367**, 234–235 (2020).
2. P. Zhou *et al.*, *Nature* **579**, 270–273 (2020).
3. J. F. Chan *et al.*, *Lancet* **395**, 514–523 (2020).
4. K. K. To *et al.*, *Emerg. Microbes Infect.* **10**, 507–535 (2021).
5. N. G. Davies *et al.*, *Science* **372**, eabg3055 (2021).
6. L. Ulrich *et al.*, *Nature* **602**, 307–313 (2022).
7. World Health Organization, Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern (WHO, 2021); https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern.
8. World Health Organization, Enhancing response to Omicron SARS-CoV-2 variant: Technical brief and priority actions for Member States (WHO, 2022); https://www.who.int/docs/default-source/coronaviruse/2022-01-07-global-technical-brief-and-priority-action-on-omicron—corr2.pdf?sfvrsn=918b09d_20.
9. L. Lu *et al.*, *Clin Infect Dis.* ciab1041 (2021).
10. L. Liu *et al.*, *Nature* **602**, 676–681 (2022).
11. S. Cele *et al.*, *Nature* **602**, 654–656 (2022).
12. Y. Cao *et al.*, *Nature* **602**, 657–663 (2022).
13. D. Planas *et al.*, *Nature* **602**, 671–675 (2022).
14. E. Cameroni *et al.*, *Nature* **602**, 664–670 (2022).
15. J. Ai *et al.*, *Emerg. Microbes Infect.* **11**, 337–343 (2022).
16. H. Ledford, *Nature* **600**, 577–578 (2021).
17. L. Espenhain *et al.*, *Euro Surveill.* **26**, 2101146 (2021).
18. L. T. Brandal *et al.*, *Euro Surveill.* **26**, 2101147 (2021).
19. J. F. Chan *et al.*, *Clin. Infect. Dis.* **71**, 2428–2446 (2020).
20. J. F. Chan *et al.*, *Clin. Infect. Dis.* **71**, 2139–2149 (2020).
21. J. F. Chan *et al.*, *Clin Infect Dis.* ciab817 (2021).
22. C. Li *et al.*, *Clin. Infect. Dis.* ciac142 (2022).
23. A. J. Zhang *et al.*, *Clin. Infect. Dis.* **73**, e503–e512 (2021).
24. G. D. de Melo *et al.*, *Sci. Transl. Med.* **13**, eabf8396 (2021).
25. J. A. Plante *et al.*, *Nature* **592**, 116–121 (2021).
26. H. Zhao *et al.*, *Emerg. Microbes Infect.* **11**, 277–283 (2022).
27. C. Zeng *et al.*, bioRxiv 2021.12.16.472934 [Preprint] (2021); doi: 10.1101/2021.12.16.472934v1.
28. K. P. Y. Hui *et al.*, *Nature* **603**, 715–720 (2022).
29. H. Shuai *et al.*, *Nature* **603**, 693–699 (2022).
30. P. J. Halfmann *et al.*, *Nature* **603**, 687–692 (2022).
31. Imperial College London, Report 50 - Hospitalisation risk for Omicron cases in England (ICL, 2021); https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-50-severity-omicron/.
32. H. Nishiura *et al.*, *J. Clin. Med.* **11**, 30 (2021).
33. World Health Organization, Enhancing readiness for Omicron (B.1.1.529): Technical brief and priority actions for member states (WHO, 2021); https://www.who.int/docs/default-source/coronaviruse/20211217-global-technical-brief-and-priority-action-on-omicron_latest-2.pdf?sfvrsn=bdd8297c_9&download=true.
34. R. Hirose *et al.*, bioRxiv 2022.01.18.476607v2 [Preprint] (2022).
35. B. J. Gardner, A. M. Kilpatrick, medRxiv [Preprint] 2021.12.10.21267594v2 (2021).

ACKNOWLEDGMENTS

We thank the staff at the Centre for Comparative Medicine Research of The University of Hong Kong for their facilitation of this study.
Funding: This study was partly supported by funding from the Health and Medical Research Fund (20190572 and COVID1903010, Projects 6, 7, and 15), the Food and Health Bureau, The Government of the Hong Kong Special Administrative Region (to S.Y., H. Chen., and J.F.-W.C.); the Collaborative Research Fund (C7060-21G) (to J.F.-W.C.) and Theme-Based Research Scheme (T11-709/21-N) (to D.-Y.J.), the Research Grants Council of the

Hong Kong Special Administrative Region; Health@InnoHK, Innovation and Technology Commission, the Government of the Hong Kong Special Administrative Region (to K.-Y.Y.); the Consultancy Service for Enhancing Laboratory Surveillance of Emerging Infectious Diseases and Research Capability on Antimicrobial Resistance for Department of Health of the Hong Kong Special Administrative Region Government (to K.-Y.Y. and J.F.-W.C.); the National Program on Key Research Project of China (2020YFA0707500 and 2020YFA0707504) (to J.F.-W.C.); Sanming Project of Medicine in Shenzhen, China (SZSM201911014) (to K.-Y.Y. and J.F.-W.C.); the High Level-Hospital Program, Health Commission of Guangdong Province, China (to K.-Y.Y. and J.F.-W.C.); Emergency Collaborative Project (EKPG22-01) of Guangzhou Laboratory (to K.K.-W.T., H. Chen, D.-Y.J., K.-Y.Y., and J.-F.W.C.); Emergency COVID-19 Project (2021YFC0866100), Major Projects on Public Security, National Key Research and Development Program (to K.K.-W.T., H. Chen, D.-Y.J., K.-Y.Y., and J.-F.W.C.); the Major Science and Technology Program of Hainan Province (ZDKJ202003) (to G.L.); the research project of Hainan Academician Innovation Platform (YSPTZX202004) (to G.L., K.-Y.Y., and J.F-.W.C.); the Hainan Talent Development Project (SRC200003) (to G.L.); the University of Hong Kong Outstanding Young Researcher Award (to J.-F.W.C.); and the University of Hong Kong Research Output Prize (Li Ka Shing Faculty of Medicine) (to J.-F.W.C.); and donations from the Shaw Foundation Hong Kong, Richard Yu and Carol Yu, Michael Seak-Kan Tong, May Tam Mak Mei Yin, Lee Wan Keung Charity Foundation Limited, Providence Foundation Limited (in memory of the late Lui Hac-Minh), Hong Kong Sanatorium and Hospital, Hui Ming, Hui Hoy and Chow Sin Lan Charity Fund Limited, The Chen Wai Wai Vivien Foundation Limited, Chan Yin Chuen Memorial Charitable Foundation, Marina Man-Wai Lee, the Hong Kong Hainan Commercial Association South China Microbiology Research Fund, the Jessie and George Ho Charitable Foundation, Perfect Shape Medical Limited, Kai Chong Tong, Tse Kam Ming Laurence, Foo Oi Foundation Limited, Betty Hing-Chu Lee, Ping Cham So, and Lo Ying Shek Chi Wai Foundation. The funding sources had no role in the study design, data collection, analysis, interpretation, or writing of the report. **Author contributions:** S.Y. and J.F.-W.C. conceived and designed the study. S.Y., Z.W.Y., R.L., K.T., A.Z., G.L., C.P.O., V.K.-M.P., C.C.-S.C., B.W.-Y.M., Z.Q., Y.X., A.W.-H.C., W.-M.C., J.D.I., H.S., J.O.-L.T., T.T.-T.Y., K.K.-H.C., C.C.Y.C., J.-P.C., C.L., C.C.-Y.Y., L.L., K.K.-W.T., H. Chen, H. Chu, and J.F.-W.C. designed and/or performed experiments. S.Y., G.L., K.K.-W.T., H. Chen, D.-Y.J., K.-Y.Y., and J.F.-W.C. acquired funding. S.Y., D.-Y.J., K.-Y.Y., and J.F.-W.C supervised the study. S.Y. and J.F.-W.C. wrote the manuscript, and all authors reviewed and edited the paper. **Competing interests:** J.F.-W.C. has received travel grants from Pfizer Corporation Hong Kong and Astellas Pharma Hong Kong Corporation Limited and was an invited speaker for Gilead Sciences Hong Kong Limited and Luminex Corporation. K.K.-W.T., H. Chen, and K.-Y.Y. have collaboration with Sinovac Biotech Ltd. and China National Pharmaceutical Group Co., Ltd. (Sinopharm). H. Chen and K.-Y.Y. have patent applications on intranasal vaccines. The other authors declare no competing interests. **Data and materials availability:** Complete sequences of the SARS-CoV-2 Omicron (hCoV-19/Hong_Kong/HKU-211129-001/2021; EPI_ISL_6841980) and Delta (hCoV-19/Hong Kong/HKU-210804-001/2021; EPI_ISL_3221329) variants are available through GISAID. All other data are provided in the manuscript or supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn8939
Materials and Methods
Figs. S1 to S10
References (*36*–*45*)
MDAR Reproducibility Checklist

Submitted 28 December 2021; accepted 17 June 2022
Published online 23 June 2022
10.1126/science.abn8939

SEMICONDUCTORS

High ambipolar mobility in cubic boron arsenide revealed by transient reflectivity microscopy

Shuai Yue[1,2,3,4]†, Fei Tian[5,6]†, Xinyu Sui[1,4], Mohammadjavad Mohebinia[7], Xianxin Wu[1,4], Tian Tong[2], Zhiming Wang[3], Bo Wu[8], Qing Zhang[9], Zhifeng Ren[5]*, Jiming Bao[2,5,7]*, Xinfeng Liu[1,4]*

Semiconducting cubic boron arsenide (c-BAs) has been predicted to have carrier mobility of 1400 square centimeters per volt-second for electrons and 2100 square centimeters per volt-second for holes at room temperature. Using pump-probe transient reflectivity microscopy, we monitored the diffusion of photoexcited carriers in single-crystal c-BAs to obtain their mobility. With near-bandgap 600-nanometer pump pulses, we found a high ambipolar mobility of 1550 ± 120 square centimeters per volt-second, in good agreement with theoretical prediction. Additional experiments with 400-nanometer pumps on the same spot revealed a mobility of >3000 square centimeters per volt-second, which we attribute to hot electrons. The observation of high carrier mobility, in conjunction with high thermal conductivity, enables an enormous number of device applications for c-BAs in high-performance electronics and optoelectronics.

In 2018, the predicted high room-temperature thermal conductivity (κ) of cubic boron arsenide (c-BAs), >1300 W m^{-1} K^{-1}, was experimentally demonstrated (*1*–*3*). At about the same time, c-BAs was also predicted to have high carrier mobility values of 1400 cm^2 V^{-1} s^{-1} for electrons and 2100 cm^2 V^{-1} s^{-1} for holes (*4*). A higher hole mobility of >3000 cm^2 V^{-1} s^{-1} was later predicted under a small 1% strain (*5*). Such a high carrier mobility is due to a weak electron–phonon interaction and small effective mass (*4*–*7*). Like those predicting the thermal conductivity of c-BAs, these calculations were based on nondefective c-BAs with high crystal quality and a very low impurity level (*4*, *5*). The simultaneous high thermal conductivity and carrier mobility makes c-BAs a promising material for many applications in electronics and optoelectronics. Despite this potential, the high mobility has not been experimentally verified (*8*). In this study, using ultrafast spatial-temporal transient reflectivity microscopy, we observed an ambipolar mobility of ~1550 cm^2 V^{-1} s^{-1} and obtained a >3000 cm^2 V^{-1} s^{-1} mobility for photoexcited hot carriers. We used photoluminescence and Raman spectroscopy to probe the relative level of p-type doping and found that a high hole concentration will substantially reduce the ambipolar mobility.

We grew c-BAs single crystals using the same seeded chemical vapor transport technique reported previously (*3*, *9*). These crystals typically appear as slabs with (111) top and bottom surfaces. We used scanning electron microscopy to image a corner facet (111) of an as-grown c-BAs slab that we labeled sample 1 (Fig. 1A). This facet is one of the eight equivalent (111) surfaces, and we chose it for mobility measurement because of its relatively high quality, which can be seen from sharp (0.02°) characteristic peaks in the x-ray diffraction (XRD) pattern (Fig. 1B and inset), a narrow (0.6 cm^{-1}) longitudinal optical (LO) phonon peak at 700 cm^{-1} in the Raman spectrum (Fig. 1C and inset) (*1*, *2*), and the characteristic bandgap photoluminescence (PL) peak at 720 nm in the PL spectrum (Fig. 1D) (*10*), indicating high-quality crystal lattices, a low mass disorder (*11*), and a low defect density, respectively (*10*). PL mapping shown in the inset of Fig. 1D also indicates the uniform crystal quality on the (111) surface (*10*). We performed all measurements at room temperature and further characterized sample 1 and a second sample, labeled sample 2 (*12*) (fig. S1).

The Hall effect is the most common technique used to measure carrier mobility, but it requires four electrical contacts on a relatively large and uniform sample. To accommodate the requirements of mobility measurement in a small sample size or in inhomogeneous materials, ultrafast pump-probe techniques have been used to perform noncontact measurements with high spatial resolution (*13*–*17*).

[1]Chinese Academy of Sciences (CAS) Key Laboratory of Standardization and Measurement for Nanotechnology, National Center for Nanoscience and Technology, Beijing 100190, China. [2]Department of Electrical and Computer Engineering and Texas Center for Superconductivity at the University of Houston (TcSUH), University of Houston, Houston, TX 77204, USA. [3]Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, China. [4]School of Nanoscience and Technology, University of Chinese Academy of Sciences, Beijing 100049, China. [5]Department of Physics and Texas Center for Superconductivity at the University of Houston (TcSUH), University of Houston, Houston, TX 77204, USA. [6]School of Materials Science and Engineering, Sun Yat-sen University, Guangzhou, Guangdong 510006, China. [7]Materials Science and Engineering Program, University of Houston, Houston, TX 77204, USA. [8]Guangdong Provincial Key Laboratory of Optical Information Materials and Technology and Institute of Electronic Paper Displays, South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou 510006, China. [9]School of Materials Science and Engineering, Peking University, Beijing 100871 China.
***Corresponding author. Email: zren@uh.edu (Z.R.); jbao@uh.edu (J.B.); liuxf@nanoctr.cn (X.L.)**
†These authors contributed equally to this work.

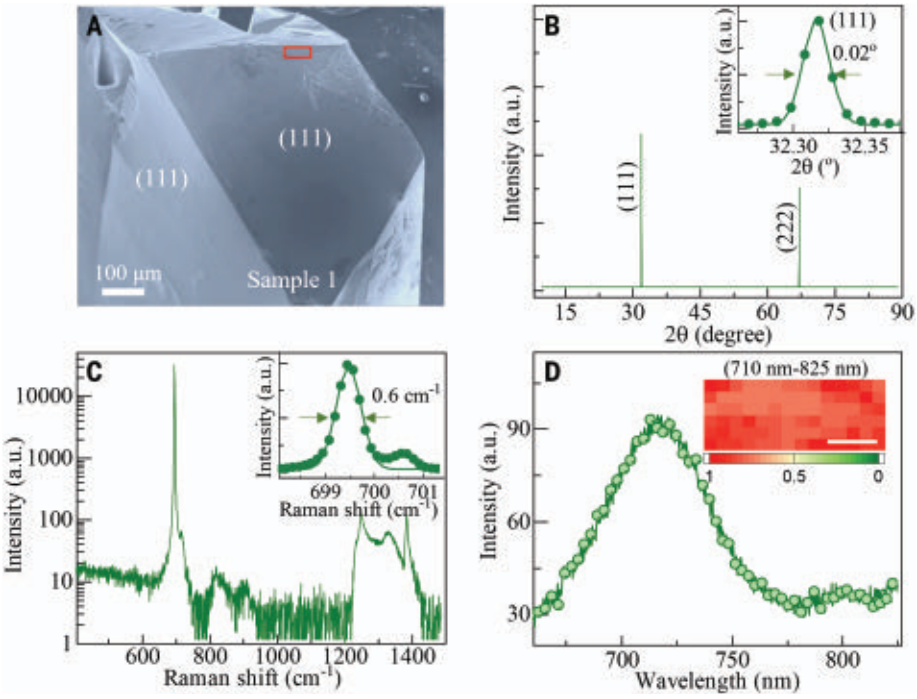


Fig. 1. Characterizations of a c-BAs single crystal (sample 1) on a corner (111) facet. (**A**) Scanning electron microscopy image. Scale bar: 100 μm. (**B**) X-ray diffraction pattern. (Inset) Magnified view of the (111) peak. (**C**) Raman spectrum excited by a 532-nm laser. (Inset) High-resolution spectrum of the LO phonon. (**D**) Photoluminescence spectrum excited by a 593-nm laser. (Inset) photoluminescence mapping from the region marked by a red rectangle in (A). Scale bar: 10 μm. a.u., arbitrary units.

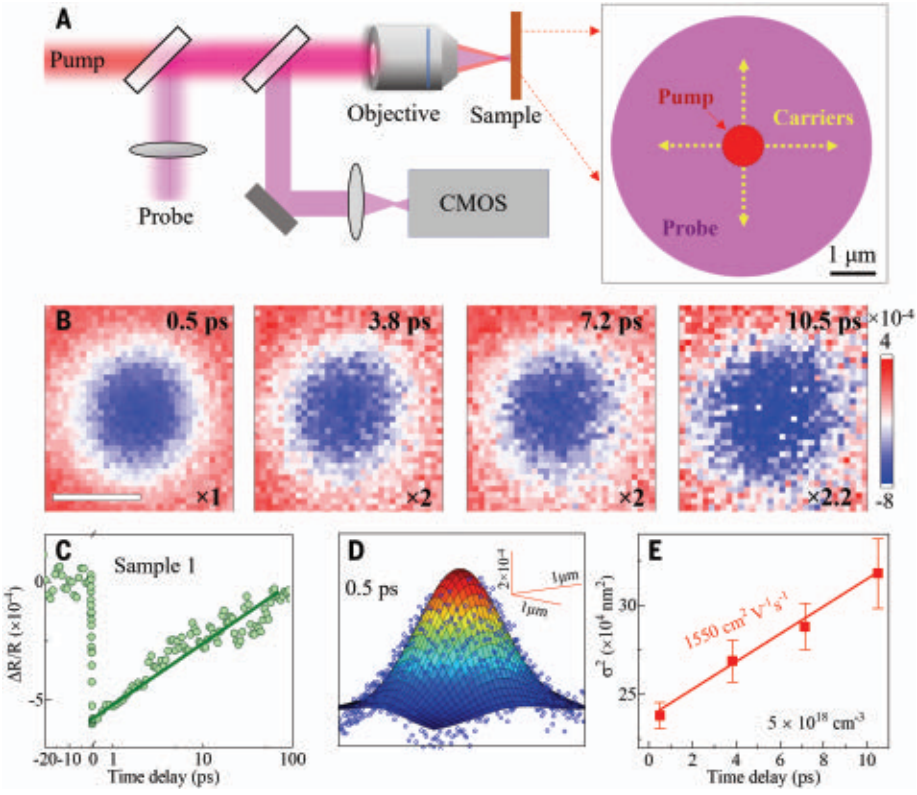


Fig. 2. Pump-probe transient reflectivity microscopy, carrier dynamics, and diffusion in sample 1. (**A**) Schematic illustration of the experimental setup. CMOS, complementary metal-oxide semiconductor. (**B**) Evolution of a 2D transient reflectivity microscopy image from a spot on sample 1. Scale bar: 1 μm. (**C**) Typical transient reflectivity dynamics (photoexcited carrier density of 5×10^{18} cm^{-3}). (**D**) Spatial profile (dots) and Gaussian fit at 0.5 ps time delay from (B) (fig. S4). (**E**) Evolution of variance of Gaussian distributions extracted from Gaussian fitting in (D). The corresponding mobility is included.

Because of our relatively thick samples, we used reflectivity rather than transmission. We focused a femtosecond pump pulse on c-BAs to photoexcite electrons and holes and monitored the diffusion of excited carriers in space and time with a time-delayed probe pulse defocused on a larger area (6 μm in diameter) (*12*) (Fig. 2A and fig. S2). We subsequently obtained an ambipolar mobility from the diffusion coefficient, D, through the Einstein relation, $D/k_BT = \mu/e$, where k_B is the Boltzmann constant, T is the temperature, μ is the mobility, and e is the elementary charge. Ambipolar mobility is given by $\mu_a = 2\mu_e\mu_h/(\mu_e + \mu_h)$, where μ_e and μ_h are the electron and hole mobility values, respectively. Because c-BAs has an electronic band structure similar to that of silicon, with an indirect bandgap in the range of 1.82 to 2.02 eV (*6*, *7*, *10*, *18*), we chose a 600-nm pump pulse and an 800-nm probe pulse to avoid the generation of hot carriers. Two-dimensional

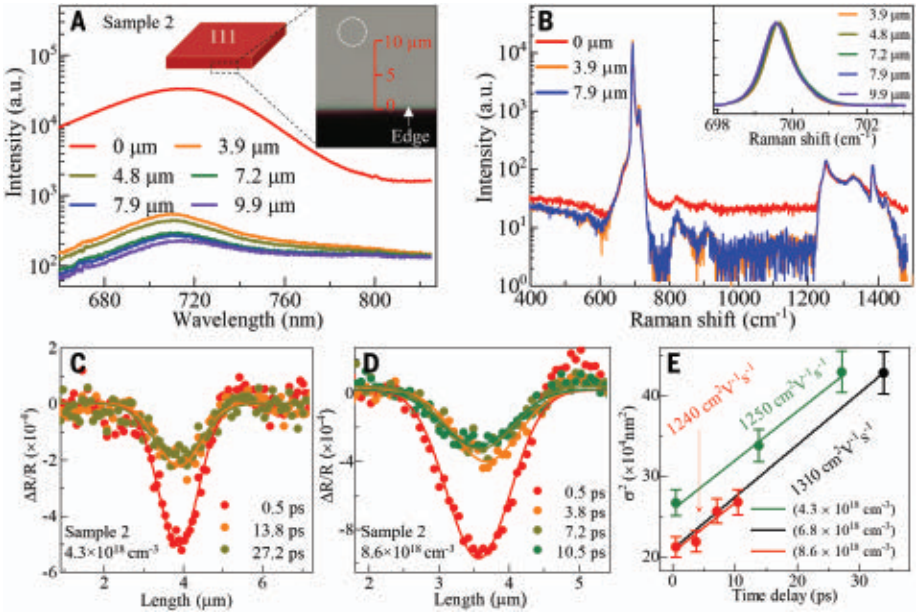


Fig. 3. Carrier diffusion on a cross-sectional surface of sample 2. (**A**) PL spectra of six locations on a cross-sectional surface with increasing distance from the edge. PL of the spot at 0 μm was taken from the (111) surface around the edge. (Inset) Optical image of the sidewall. Dashed circle indicates location for pump-probe measurements in (C) to (E). (**B**) Raman spectra of three of the six locations shown in (A). (Inset) Magnified view of the phonon line in the spectra of the five sidewall locations. (**C** and **D**) Spatial profiles (dots) and Gaussian fits (curves) of photoexcited carriers at initial concentrations of 4.3×10^{18} cm^{-3} and 8.6×10^{18} cm^{-3}, respectively, from a location indicated by the dashed circle in (A). (**E**) Variance and ambipolar mobility values from (C), (D), and fig. S6.

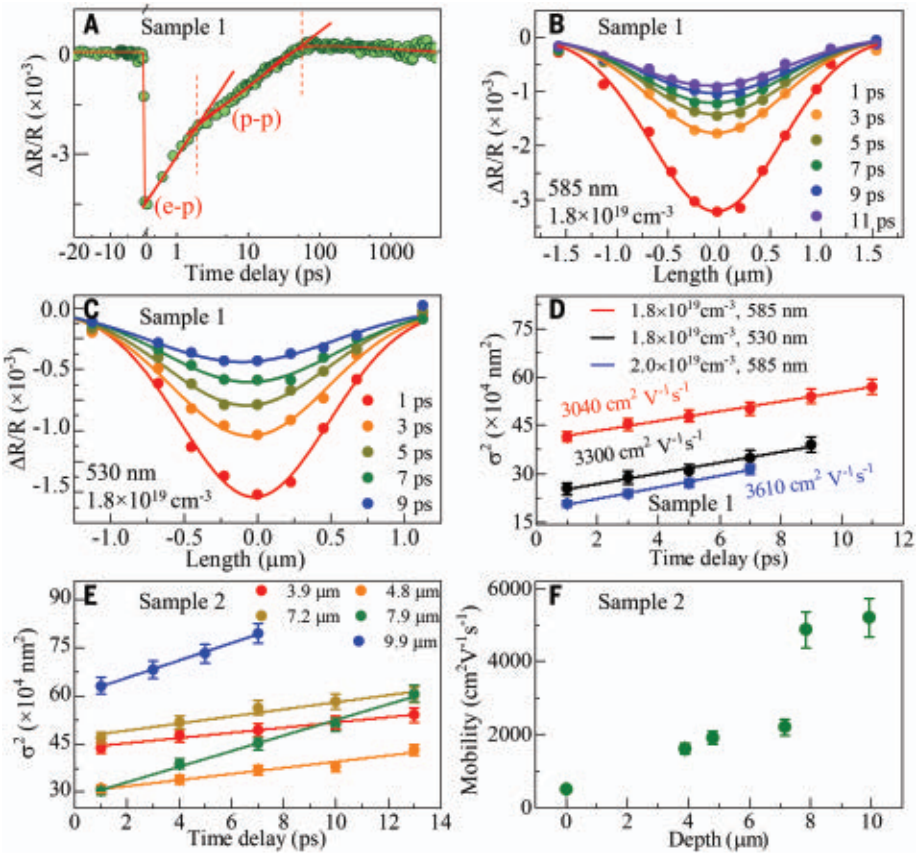


Fig. 4. Transient reflectivity microscopy and carrier diffusion measured using a 400-nm pump and a 585- or 530-nm probe. (**A**) Representative pump-probe transient reflectivity curve from sample 1. The probe wavelength is 585 nm. (**B** and **C**) Spatial profiles (dots) and Gaussian fits (curves) of transient reflectivity from a spot in sample 1 measured using 585- and 530-nm probes, respectively. (**D**) Evolution of the variances of carrier density distributions and carrier mobility from (B), (C), and fig. S10. (**E** and **F**) Variance and ambipolar mobility results, respectively, for sample 2 at six locations corresponding to those shown in Fig. 3, A and B.

(2D) diffusion images in Fig. 2B show the expansion of carriers over 10 ps, and a representative time-resolved reflectivity as a function of the time delay between the pump and the probe is shown in Fig. 2C. A sudden negative differential reflectivity indicates a dominant electronic contribution, because reflectivity increases with lattice temperature (*12*, *19*) (fig. S3).

The spread of distributions in Fig. 2B reflects diffusion of photoexcited electrons and holes in space and time, and they can be well fit by Gaussian functions (Fig. 2D). The change in the variance σ^2 of carrier distributions is plotted in Fig. 2E. The linear increase in the variance with increasing time delay is a signature of diffusion, and the diffusion coefficient, D, can be calculated from the slope using the equation $\sigma_t^2 = \sigma_0^2 + \alpha Dt$, where α is a constant depending on the dimensions of the system and detection configuration (*15*). We chose an α of 2 for our experiment because of the much larger laser penetration (excitation) depth (60 μm at 600 nm) compared with the thin top layer sampled by the

probe beam [20 nm at 800 nm, given by $\lambda/4\pi n$ (*13*, *15*–*17*, *19*), where n is the refractive index of c-BAs (*18*)]. From the slope of the curve shown in Fig. 2E and the Einstein relation, $D/k_BT = \mu/e$, we obtained an ambipolar diffusion coefficient of ~39 $cm^2\ s^{-1}$ and an ambipolar mobility of 1550 ± 120 $cm^2\ V^{-1}\ s^{-1}$, close to the predicted value (*4*).

Given that the properties of c-BAs are not uniform even within a single crystal, especially in the direction perpendicular to (111) surfaces (*3*, *10*), we tested a cross-sectional surface of a relatively thin (30-μm-thick) crystal labeled sample 2 (*12*) (fig. S1). An optical image of the sample 2 sidewall is shown in the inset of Fig. 3A. We obtained PL spectra from several spots at different distances from the edge (Fig. 3A). The PL intensity increases with decreasing distance from the edge and exhibits a noticeable jump upon reaching the (111) surface, which agrees with our previous finding of a drastic change in PL from one surface to the opposite surface of a single crystal slab (*10*). Corresponding Raman spectra from the same locations are shown in Fig. 3B and its inset. Similar to the PL results, the Raman spectrum of the (111) surface differs substantially from those of the sidewall. We chose a spot ~11 μm from the edge (Fig. 3A, dashed circle in inset) and used three pump fluences to create different carrier densities, the reflectivity distributions of which are shown in Fig. 3, C and D, and fig. S6 (*12*). We plotted the evolution of the variances and obtained an ambipolar mobility of ~1300 $cm^2\ V^{-1}\ s^{-1}$ (Fig. 3E), indicating the negligible effect of carrier density on the mobility of sample 2 owing to nonlinear effects such as Auger recombination.

The high carrier mobility of c-BAs is enabled by its distinctive weak electron–phonon interaction and its phonon–phonon scattering, which should also enable the generation of high-mobility hot carriers (*20*). To prove this, we used a 400-nm pulse as a pump and selected a particular band (585 or 530 nm) with an optical filter from a white light continuum beam as a probe pulse (*12*) (fig. S7). A typical transient reflectivity curve of a probe (585 nm) from sample 1 is shown in Fig. 4A. In contrast to the single exponential decay previously observed when excited by a 600-nm pump (Fig. 2C), the dynamics of photoexcited carriers excited by the 400-nm pump consist of three exponential decays: a fast exponential decay with a ~1-ps lifetime, a slow decay of ~20 ps, and an even slower decay on the order of 1 ns (*21*). These decays correspond to rapid relaxation of high-energy photoexcited carriers, further relaxation of carriers to the conduction and valence band edges, and a combination of lattice heating and recombination and trapping of electrons and holes at the band edges, respectively (*20*, *21*), in good agreement with the theoretical prediction (*20*). To obtain the diffusion coefficient of the carriers in sample 1, we used a simpler method by varying the relative displacement between focused pump and probe beams along one direction (*12*) (figs. S7 and S8). We plotted the resulting spatial profiles of the reflectivity after 1 ps for probe wavelengths of 530 and 585 nm (Fig. 4, B and C) (*12*) and obtained an ambipolar diffusion coefficient of 80 $cm^2\ s^{-1}$ and an ambipolar mobility of ~3200 $cm^2\ V^{-1}\ s^{-1}$ (Fig. 4D). Mobility of ~3600 $cm^2\ V^{-1}\ s^{-1}$ was obtained from the same spot as that shown in Fig. 2 for sample 1. These values are much larger than the predicted ambipolar mobility of 1680 $cm^2\ V^{-1}\ s^{-1}$ (*4*).

Using the same 400-nm pump, we also measured the ambipolar mobility of sample 2 at six locations corresponding to those shown in Fig. 3, A and B. The evolution of variance of carrier distribution at these spots is shown in Fig. 4, E and F, and fig. S11. The differences in the initial values of the variances at 1 ps are due to the different spot sizes of the pump and probe beams in each measurement. The mobility clearly changes drastically across the sidewall, with the highest mobility (5200 ± 600 $cm^2\ V^{-1}\ s^{-1}$) observed at a depth of 9.9 μm. Although local strain could result in such prominent carrier mobility enhancement (*5*), we did not see any noticeable Raman shift among these locations (Fig. 3B). We thus attribute the high ambipolar mobility to photoexcited hot carriers, which exhibit high carrier diffusion coefficient and mobility values (*20*, *22*–*24*).

The position-dependent mobility on the sidewall of sample 2 reveals that p-type doping in c-BAs can substantially reduce its mobility. Heavy p-type doping on the (111) surface can be seen from the Fano line shape of the LO phonon at 700 cm^{-1} and the higher background level around 1000 cm^{-1} (Fig. 3B) (*2*, *8*). This gradually increased doping level toward the (111) surface is further supported by the corresponding increased PL intensity (*10*, *25*). P-type doping will result in reduced carrier mobility owing to the presence of ionized dopants (these dopants are already activated) and a lower electron mobility than hole mobility, because minority carriers will dominate the carrier dynamics. The latter is supported by our observation of a higher ambipolar mobility in p-type silicon than in undoped silicon (*12*) (figs. S12 and S13). Clearly, the enhanced PL intensity observed in the c-BAs samples in the current study indicates that p-type doping has only introduced shallow acceptors rather than nonradiative deep levels (*10*, *25*). Because hot carriers can also be generated by electrical injection and low-intensity light, both hot carriers and fully relaxed carriers can be used for high-speed optoelectronic devices and high-efficiency solar cells in conjunction with the high mobility of the band-edge carriers.

REFERENCES AND NOTES

1. J. S. Kang, M. Li, H. Wu, H. Nguyen, Y. Hu, *Science* **361**, 575–578 (2018).
2. S. Li *et al*., *Science* **361**, 579–581 (2018).
3. F. Tian *et al*., *Science* **361**, 582–585 (2018).
4. T.-H. Liu *et al*., *Phys. Rev. B* **98**, 081203 (2018).
5. K. Bushick, S. Chae, Z. Deng, J. T. Heron, E. Kioupakis, *NPJ Comput. Mater.* **6**, 3 (2020).
6. J. L. Lyons *et al*., *Appl. Phys. Lett.* **113**, 251902 (2018).
7. K. Bushick, K. Mengle, N. Sanders, E. Kioupakis, *Appl. Phys. Lett.* **114**, 022101 (2019).
8. X. Chen *et al*., *Chem. Mater.* **33**, 6974–6982 (2021).
9. F. Tian *et al*., *Appl. Phys. Lett.* **112**, 031903 (2018).
10. S. Yue *et al*., *Mater. Today Phys.* **13**, 100194 (2020).
11. A. Rai, S. Li, H. L. Wu, B. Lv, D. G. Cahill, *Phys. Rev. Mater.* **5**, 013603 (2021).
12. See supplementary materials.
13. Y. Wan *et al*., *Nat. Chem.* **7**, 785–792 (2015).
14. M. M. Gabriel *et al*., *Nano Lett.* **13**, 1336–1340 (2013).
15. N. S. Ginsberg, W. A. Tisdale, *Annu. Rev. Phys. Chem.* **71**, 1–30 (2020).
16. R. Wang *et al*., *Phys. Rev. B* **86**, 045406 (2012).
17. L. Yuan *et al*., *Nat. Mater.* **19**, 617–623 (2020).
18. B. Song *et al*., *Appl. Phys. Lett.* **116**, 141903 (2020).
19. A. J. Sabbah, D. M. Riffe, *Phys. Rev. B* **66**, 165217 (2002).
20. S. Sadasivam, M. K. Y. Chan, P. Darancet, *Phys. Rev. Lett.* **119**, 136602 (2017).
21. Z. Y. Tian *et al*., *Phys. Rev. B* **105**, 174306 (2022).
22. A. Block *et al*., *Sci. Adv.* **5**, eaav8965 (2019).
23. B. A. Ruzicka *et al*., *Phys. Rev. B* **82**, 195414 (2010).
24. E. Najafi, V. Ivanov, A. Zewail, M. Bernardi, *Nat. Commun.* **8**, 15177 (2017).
25. S. Y. Lim *et al*., *IEEE J. Photovolt.* **3**, 649–655 (2013).

ACKNOWLEDGMENTS

We thank X. Bo and L. Yang for help with high-resolution Raman and XRD measurements. We thank J. Ding and X. Qiu for help with the AFM measurement. We thank J. Zhao for helpful discussion. **Funding:** The work performed in China was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB36000000); the Ministry of Science and Technology (2017YFA0205004); the National Natural Science Foundation of China (22173025, 22073022, 11874130, and 52172171); the CAS Instrument Development Project (Y950291); and the DNL Cooperation Fund, CAS (DNL202016). The work performed at the University of Houston (UH) is supported by the Office of Naval Research under Multidisciplinary University Research Initiative grant N00014-16-1-2436, the Welch Foundation (E-1728), and a UH Small Equipment Grant (000182016). **Author contributions:** X.L., J.B., and Z.R. conceived of the project. S.Y. performed the Raman, PL, transient reflectivity, and transient reflectivity microscopy experiments. F.T. grew the crystal samples. X.S. and S.Y. built the transient reflectivity setup. S.Y. built the transient reflectivity microscopy setup. M.M. performed the carrier diffusion and laser heating simulation. X.W. performed the high-resolution Raman experiment. T.T. performed the temperature-dependent reflectivity experiment. B.W. performed the carrier diffusion theoretical analysis. J.B., S.Y., Z.R., X.L., Z.W., and Q.Z. wrote the paper. All authors contributed to the discussion of the results and writing of the manuscript. **Competing interests:** The authors declare that they have no competing interests. The c-BAs crystals were grown by the method disclosed in US patent publication 20210269318 and a new patent filing on high mobility. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn4727
Materials and Methods
Supplementary Text
Figs. S1 to S13
References (*26*, *27*)

Submitted 1 December 2021; accepted 16 June 2022
10.1126/science.abn4727

SEMICONDUCTORS

High ambipolar mobility in cubic boron arsenide

Jungwoo Shin[1]†, Geethal Amila Gamage[2]†, Zhiwei Ding[1]†, Ke Chen[1], Fei Tian[2], Xin Qian[1], Jiawei Zhou[1], Hwijong Lee[3], Jianshi Zhou[3], Li Shi[3], Thanh Nguyen[4], Fei Han[4], Mingda Li[4], David Broido[5], Aaron Schmidt[1], Zhifeng Ren[2]*, Gang Chen[1]*

Semiconductors with high thermal conductivity and electron-hole mobility are of great importance for electronic and photonic devices as well as for fundamental studies. Among the ultrahigh–thermal conductivity materials, cubic boron arsenide (c-BAs) is predicted to exhibit simultaneously high electron and hole mobilities of >1000 centimeters squared per volt per second. Using the optical transient grating technique, we experimentally measured thermal conductivity of 1200 watts per meter per kelvin and ambipolar mobility of 1600 centimeters squared per volt per second at the same locations on c-BAs samples at room temperature despite spatial variations. Ab initio calculations show that lowering ionized and neutral impurity concentrations is key to achieving high mobility and high thermal conductivity, respectively. The high ambipolar mobilities combined with the ultrahigh thermal conductivity make c-BAs a promising candidate for next-generation electronics.

The performance of microelectronic and optoelectronic devices benefits from semiconductors with simultaneously high electron and hole mobilities and high thermal conductivity (*1*, *2*). However, mobility and thermal conductivity measurements have thus far identified no such materials. Two of the most widely used semiconductors, silicon and gallium arsenide (GaAs), for example, have high room temperature (RT) electron mobilities of μ_e = 1400 $cm^2V^{-1}s^{-1}$ and 8500 $cm^2V^{-1}s^{-1}$, respectively. However, their corresponding RT hole mobilities (μ_h = 450 $cm^2V^{-1}s^{-1}$ for Si and 400 $cm^2V^{-1}s^{-1}$ for GaAs) and thermal conductivities (κ_{RT} = 140 $Wm^{-1}K^{-1}$ for Si and 45 $Wm^{-1}K^{-1}$ for GaAs) are lower than desired. Although graphene has high electron and hole mobilities and a high in-plane thermal conductivity, the cross-plane heat conduction is low (*3*, *4*). Diamond has the highest RT thermal conductivity and excellent electron and hole mobilities; however, its large bandgap of 5.4 eV hinders its effective doping and utilization as a semiconductor material (*5*). Recently, first-principles calculations have predicted that cubic boron arsenide (c-BAs) should have exceptionally high RT thermal conductivity of ~1400 $Wm^{-1}K^{-1}$, 10 times as high as that of Si. This high value stems from its unusual phonon dispersions and chemical bonding properties that promote simultaneously weak three-phonon and four-phonon scattering (*6*–*8*). This prediction has now been demonstrated experimentally (*9*–*11*), with measured c-BAs thermal conductivities in the range of κ_{RT} = 1000 to 1300 $Wm^{-1}K^{-1}$, identifying c-BAs as the most thermally conductive semiconductor other than diamond.

First-principles calculations have also predicted that c-BAs should have simultaneously high RT electron and hole mobilities of μ_e = 1400 $cm^2V^{-1}s^{-1}$ and μ_h = 2100 $cm^2V^{-1}s^{-1}$, respectively (*12*). The major reason for such high electron and hole mobilities is the high energy and low occupation of polar optical phonons in c-BAs, which give rise to weak carrier scattering. This feature distinguishes c-BAs from other III-V semiconductors, which have high electron mobility but much lower hole mobility, where μ_e/μ_h > 10 to ~100 (*13*, *14*), except for AlSb (μ_e = 200 $cm^2V^{-1}s^{-1}$ and μ_h = 400 $cm^2V^{-1}s^{-1}$).

Despite the promising theoretical predictions, experimental measurements have not found high mobilities in BAs. Similar to the history of the development of other III-V semiconductors (*15*), the initial quality of c-BAs crystals has been limited by large and nonuniform defect concentrations. Because traditional bulk transport measurement methods can only obtain the defect-limited behaviors instead of the intrinsic properties, the high defect densities in c-BAs crystals have prevented such measurements from assessing the validity of the predicted high mobilities. Furthermore, previous studies have shown that thermal conductivity and electronic mobility do not seem to have a strong relationship with each other. Kim *et al.* measured κ_{RT} = 186 $Wm^{-1}K^{-1}$ and estimated μ_h = 400 $cm^2V^{-1}s^{-1}$ of a c-BAs microrod sample (*16*). Chen *et al.* measured κ_{RT} = 920 $Wm^{-1}K^{-1}$ and μ_h = 22 $cm^2V^{-1}s^{-1}$ of millimeter-scale c-BAs crystals (*17*). The obtained mobilities are much lower than the calculated mobility and do not show a clear correlation with the measured thermal conductivity. The origins of (i) the discrepancy between ab initio calculations and experiments and (ii) the decoupling between thermal and electrical properties have not been identified.

We used an optical transient grating (TG) method to measure electrical mobility and thermal conductivity on the same spot of c-BAs single crystals. Our experiments confirm that c-BAs has simultaneous high thermal conductivity and high electron and hole mobilities. Using ab initio calculations, we show that ionized impurities strongly scatter charge carriers, whereas neutral impurities are mainly responsible for the thermal conductivity reduction. These findings establish c-BAs as the only known semiconductor with this combination of desirable properties and place it among the ideal materials for next-generation microelectronics applications.

We prepared c-BAs samples using multistep chemical vapor transport with varying conditions (*18*) (figs. S1 and S2). We used scanning electron microscopy (SEM) to image a c-BAs single crystal with a thickness of ~20 μm (Fig. 1, A and B) and confirmed the cubic structure with x-ray diffraction (XRD) (Fig. 1C), in agreement with the literature (*19*).

We used photoluminescence (PL) and Raman spectroscopies to identify the nonuniform impurity distribution in c-BAs (*17*, *20*). We measured the PL spectrum (Fig. 1D) and performed two-dimensional (2D) PL mapping of c-BAs crystals (Fig. 1E). Local bright spots indicate the spatial differences in charge carrier density and recombination dynamics. We also measured the Raman spectrum (Fig. 1F) and performed 2D Raman background scattering intensity (I_{BG}) mapping (Fig. 1G). The strong Raman peak at ~700 cm^{-1} is associated with the longitudinal optical (LO) mode of c-BAs at the zone center. The full width at half maximum of the LO peak and I_{BG} can be attributed to mass disorder resulting from impurities, responsible for large κ variation (*11*, *21*).

We used the TG technique (*22*–*24*) (Fig. 2A) to simultaneously measure electrical and thermal transport on multiple spots (Fig. 1, circles a to d). Two femtosecond laser pulses (pump) with wavevectors $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$ create sinusoidal optical interference on the c-BAs samples, exciting electron-hole pairs accordingly (fig. S3). A third laser pulse ($\boldsymbol{k}_3$; probe) arrives at the sample spot after delay time t, which is subsequently diffracted to the direction of $\boldsymbol{k}_1 - \boldsymbol{k}_2 + \boldsymbol{k}_3$ and mixed with a fourth pulse ($\boldsymbol{k}_4$) for heterodyne detection. As the photoexcited carriers undergo diffusion and recombination, the corresponding diffraction signal decays with t. We show the calculated time-dependent electron-hole profile in c-BAs in Fig. 2B and figs. S4 and S5.

Diffusion and recombination of photoexcited carriers result in a fast exponential decay in the TG signal (t < 1 ns), followed by a slower

[1]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Department of Physics and Texas Center for Superconductivity, University of Houston, Houston, TX 77204, USA. [3]Materials Science and Engineering Program, The University of Texas at Austin, Austin, TX 78712, USA. [4]Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [5]Department of Physics, Boston College, Chestnut Hill, MA 02467, USA.
*Corresponding author. Email: zren2@central.uh.edu (Z.R.); gchen2@mit.edu (G.C.)
†These authors contributed equally to this work.

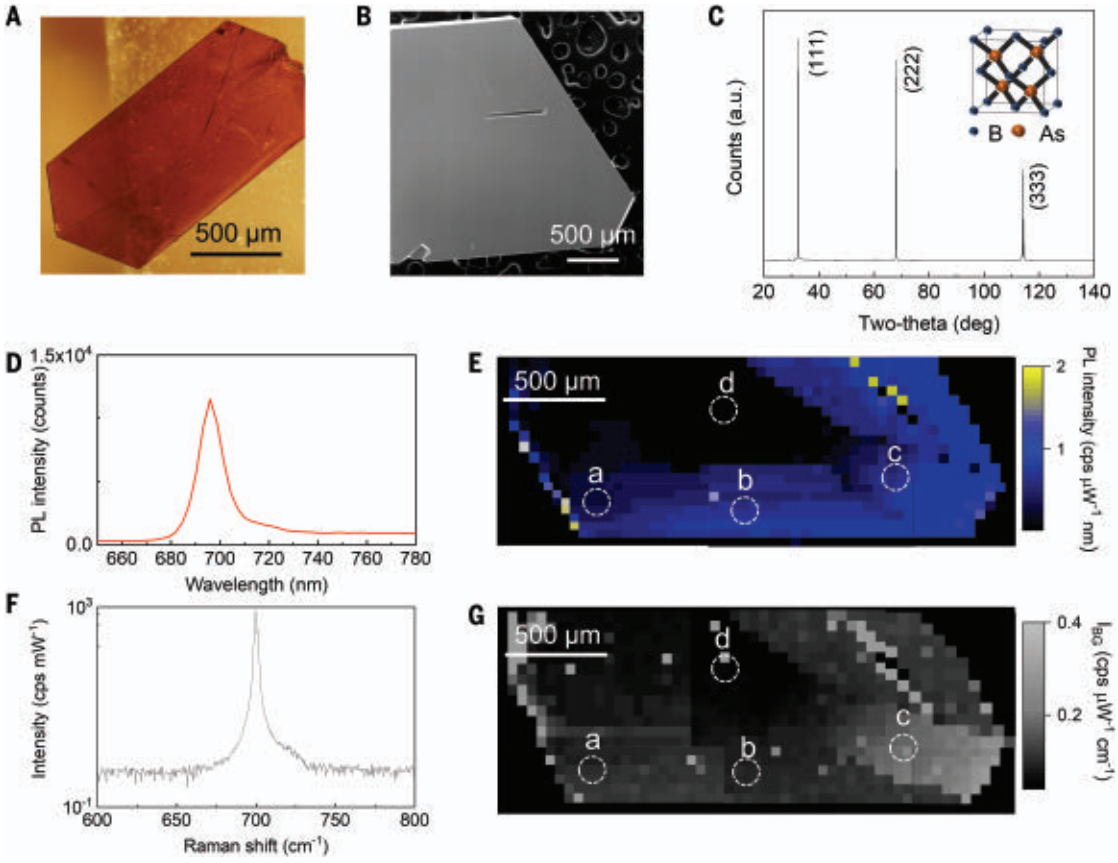


Fig. 1. Optical characterization of c-BAs single crystals. (**A**) Optical photograph. (**B**) SEM image. (**C**) XRD. a.u., arbitrary units; deg, degrees. (**D** and **E**) A typical PL spectrum (D) and 2D PL intensity mapping (E) integrated over 100-nm spectrum range for each spot. The dashed circles show TG measurement spots (a to d). cps, counts per second. (**F** and **G**) A typical Raman spectrum (F) and 2D mapping of background Raman scattering intensity (G) integrated over 100 cm^{-1} for each spot.

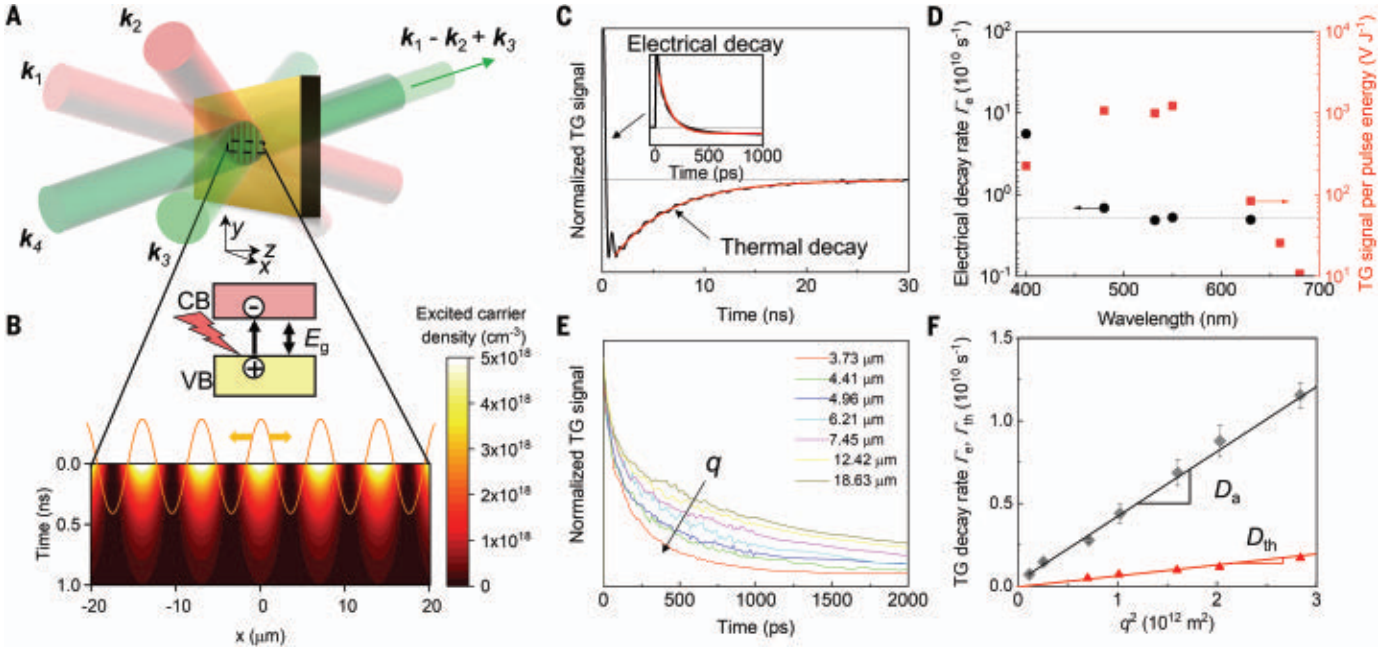


Fig. 2. Thermal and electron transport measurements. (**A**) Schematic illustration of TG experiments. (**B**) Calculated time-dependent electron-hole pair density in c-BAs. CB, conduction band; VB, valence band; E_g, bandgap. (**C**) TG signal for c-BAs. Thermal conductivity is calculated from exponential fitting (red line). (**D**) Wavelength-dependent electrical decay rate Γ_e and TG peak amplitude. (**E**) TG signal with varying diffraction grating periods q. (**F**) Electrical decay rate (Γ_e) and thermal decay rate (Γ_{th}) versus q^2. Error bars show experimental uncertainties.

thermal decay (t > 1 ns) with an opposite sign (Fig. 2C). The short and long time decays are used to calculate charge carrier mobility and thermal conductivity on the same spot, respectively (see fig. S6 for details). Thermal conductivity is directly calculated from the exponential fitting of the long time decay (red line). The electrical decay is sensitive to the wavelength of the pump pulses. We use an optical parametric amplifier (OPA) to match the wavelength of the pump beam with the bandgap (2.02 eV) of c-BAs to avoid excitation of high-energy electrons that can lead to hot electrons and holes with different scattering dynamics and mobilities (*25*). We also determined the wavelength-dependent electrical decay rate Γ_e and the lock-in amplifier amplitude of the TG peak (Fig. 2D). TG decays much faster at shorter wavelengths (λ < 500 nm) and reaches a plateau near the bandgap (λ ~ 600 nm) followed by signal loss for photon energy below the bandgap (λ > 650 nm) (fig. S7). The slopes of electrical decay Γ_e and thermal decay Γ_{th} versus q^2 (Fig. 2, E and F) are equivalent to the ambipolar diffusivity D_a and thermal diffusivity D_{th} of c-BAs. D_a is subsequently converted to ambipolar mobility $\mu_a = eD_a/k_BT = 2\mu_e\mu_h/(\mu_e + \mu_h)$, which is dominated by the low mobility carrier, where k_B is the Boltzmann constant, e is the elementary charge, and T is temperature.

We measured a wide variation of the RT κ and μ_a for spots a to d (a: 920 $Wm^{-1}K^{-1}$ and 731 $cm^2V^{-1}s^{-1}$; b: 1132 $Wm^{-1}K^{-1}$ and 1482 $cm^2V^{-1}s^{-1}$; c: 163 $Wm^{-1}K^{-1}$ and 331 $cm^2V^{-1}s^{-1}$; d: 211 $Wm^{-1}K^{-1}$ and 328 $cm^2V^{-1}s^{-1}$). This large spatial variation of thermal and electrical properties can be attributed to corresponding variations in impurity density. A higher impurity density lowers PL intensity and increases I_{BG}. To corroborate this trend, we intentionally doped c-BAs with C (batch **IV**) and measured κ = 200 to 953 $Wm^{-1}K^{-1}$ and μ_a = 195 to 416 $cm^2V^{-1}s^{-1}$ along with large variation in I_{BG} and low PL intensity (figs. S8 and S9).

Common impurities in c-BAs are group IV elements, such as C and Si. These impurities can serve as electron acceptors in c-BAs because of low formation energies (*26*). Space charges created by ionized impurities introduce distortions in the local bonding environment, driving distinct phonon scattering mechanisms. The κ of c-BAs can be calculated by solving the phonon Boltzmann transport equation, including three- and four-phonon scattering and phonon-scattering by neutral (solid lines) and charged (dashed lines) group IV impurities on B or As sites (*27*, *28*) (Fig. 3A). Our calculated κ decreases with increasing mass difference between the impurity and host atoms. Upon impurity ionization, the number of valence electrons of the impurity (IV) matches that of B or As (III or V), resulting in weaker bond perturbations than those from the neutral impurities. Consequently, the thermal conductivity reduction from ionized impurities is smaller than that caused by the un-ionized impurities, especially when the substituted impurity has a similar mass to that of the host atom—i.e., Ge_{As}^- and C_B^+.

The bond perturbation and Coulomb potential of impurities modify electron and hole transport dynamics in c-BAs differently. Building on recent developments in computing formation energies for charged impurities (*29*), we used ab initio calculations to study the effect of group IV impurities on the RT μ_a of c-BAs (Fig. 3B). We show electron-phonon scattering and long- and short-range defect scattering for holes in c-BAs with Si_{As}^- (see fig. S10 for details) (Fig. 3C). Long-range Coulombic interaction with charged impurities is found to be the dominant scattering mechanism near the band edge. The lack of a Coulomb potential for neutral impurities results in a weaker carrier scattering, causing μ_a to not decrease until the concentration approaches 10^{18} cm^{-3}, where the electron-neutral impurity scattering starts to show an effect. However, μ_a decreases markedly with charged impurities from 10^{16} cm^{-3}, regardless of the mass of the impurity.

We elucidated the different effects of neutral and charged impurities on κ and μ_a (Fig. 3D). Neutral impurities more strongly suppress κ because of stronger bond perturbations compared with those in charged impurities (*27*). Charged impurities predominantly contribute to μ_a reduction regardless of their mass as a result of Coulombic scattering. Charged impurities with masses similar to that of the host atom would exhibit κ_{RT} above 1000 $W\,m^{-1}\,K^{-1}$, even at a high impurity density of 10^{19} cm^{-3}, and μ_a is significantly reduced to below 400 $cm^2V^{-1}s^{-1}$ at a moderate level of 10^{18} cm^{-3}.

We can also highlight the contrasting trends in κ and μ_a with neutral and charged impurities from batches 0 to IV (Fig. 4A and table S1) (*18*). Solid and dashed lines in Fig. 4 show the trajectories of the calculated μ_a and κ with neutral Si_{As}^0 and charged Si_{As}^- from 10^{16} to 10^{20} cm^{-3}, respectively. Scattered points are the measured μ_a and κ values of samples from different batches, labeled with different colors. All measured data fit into the area between

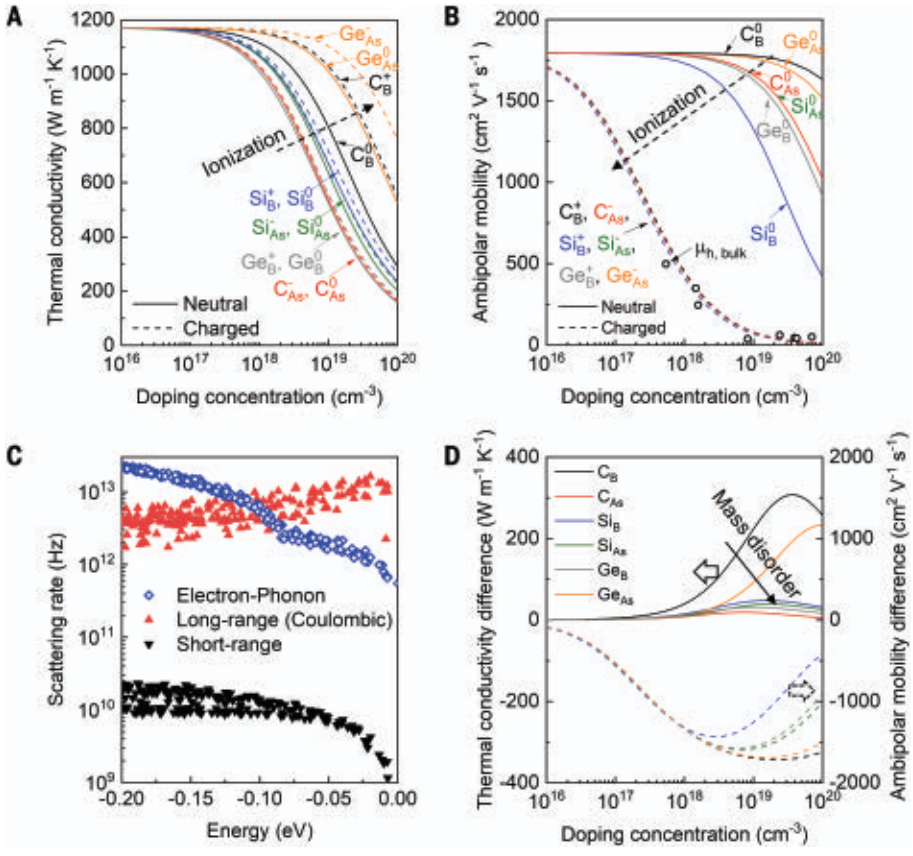


Fig. 3. Theoretical calculation of the impurity effects on thermal conductivity and mobility. (**A** and **B**) Calculated thermal conductivity (A) and ambipolar mobility (B) with neutral (solid lines) and charged (dashed lines) group IV impurities. Open circles are μ_h values of bulk samples measured by electrical probes (fig. S12). (**C**) Calculated electron-phonon and short- and long-range impurity scattering rates for holes. Zero of energy is at the valence band maximum. $\left(Si_{As}^- = 10^{18}cm^{-3}\right)$. (**D**) Thermal conductivity (solid lines) and mobility (dashed lines) differences between charged and neutral impurities.

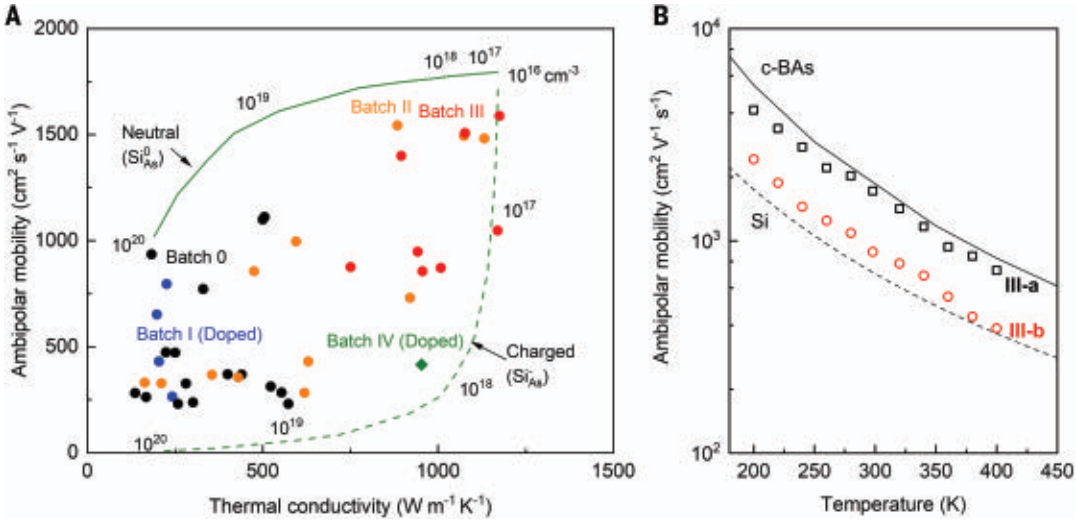


Fig. 4. Ambipolar mobility and thermal conductivity of c-BAs. (**A**) Measured mobility and thermal conductivity of c-BAs from different batches (batches **0**, **I**, **II**, **III**, and **IV**). See table S1 for details. The solid and dashed lines show the calculated μ_a and κ with varying concentrations of neutral Si^0_{As} and charged Si^-_{As}, respectively. Typical uncertainties for μ_a and κ are 11%. (**B**) Temperature-dependent ambipolar mobility of c-BAs (**III-a** and **III-b**). The solid and dashed lines show calculated μ_a of pristine c-BAs and Si, respectively (*32*).

the trajectory curves. Among the high-quality c-BAs batch (**III**), we measure μ_a = 1600 ± 170 $cm^2V^{-1}s^{-1}$ and κ = 1200 ± 130 $Wm^{-1}K^{-1}$. We measured the temperature-dependent μ_a of two different spots (**III-a** and **III-b**) of high-quality samples (fig. S11). Our measured μ_a for **III-a** shows good agreement with calculations (Fig. 4B). Hall measurements of the bulk samples provide μ_h and carrier concentration p averaged over the entire sample with spatially varied impurity concentration. The measured bulk μ_h plotted in Fig. 3B (see fig. S12 for details) is limited by the average impurity concentrations rather than local spots with low impurities.

The high–spatial resolution TG measurements provide clear evidence of simultaneously high electron and hole mobilities in c-BAs and demonstrate that through the elimination of defects and impurities, c-BAs could exhibit both high thermal conductivity and high electron and hole mobilities. Additionally, the observed weak correlation between the local thermal conductivity and mobility is caused by the different effects that neutral and ionized impurities have on these quantities. This notable combination of electronic and thermal properties, along with a thermal expansion coefficient and lattice constant that are closely matched to common semiconductors such as Si and GaAs (*30*, *31*), make c-BAs a promising material for integrating with current and future semiconductor manufacturing processes and addressing the grand challenges in thermal management for next-generation electronics.

REFERENCES AND NOTES

1. X. Qian, J. Zhou, G. Chen, *Nat. Mater.* **20**, 1188–1202 (2021).
2. G. Chen, *Nat. Rev. Phys.* **3**, 555–569 (2021).
3. K. S. Novoselov *et al.*, *Science* **306**, 666–669 (2004).
4. A. A. Balandin *et al.*, *Nano Lett.* **8**, 902–907 (2008).
5. C. J. H. Wort, R. S. Balmer, *Mater. Today* **11**, 22–28 (2008).
6. L. Lindsay, D. A. Broido, T. L. Reinecke, *Phys. Rev. Lett.* **111**, 025901 (2013).
7. D. A. Broido, L. Lindsay, T. L. Reinecke, *Phys. Rev. B* **88**, 214303 (2013).
8. T. L. Feng, L. Lindsay, X. L. Ruan, *Phys. Rev. B* **96**, 161201 (2017).
9. J. S. Kang, M. Li, H. Wu, H. Nguyen, Y. Hu, *Science* **361**, 575–578 (2018).
10. F. Tian *et al.*, *Science* **361**, 582–585 (2018).
11. S. Li *et al.*, *Science* **361**, 579–581 (2018).
12. T. H. Liu *et al.*, *Phys. Rev. B* **98**, 081203 (2018).
13. D. L. Rode, *Phys. Rev. B* **3**, 3287–3299 (1971).
14. A. Nainani, B. R. Bennett, J. B. Boos, M. G. Ancona, K. C. Saraswat, *J. Appl. Phys.* **111**, 103706 (2012).
15. J. I. Pankove, T. D. Moustakas, *Semicond. Semimet.* **50**, 1–10 (1997).
16. J. Kim *et al.*, *Appl. Phys. Lett.* **108**, 201905 (2016).
17. X. Chen *et al.*, *Chem. Mater.* **33**, 6974–6982 (2021).
18. Materials and methods are available as supplementary materials online.
19. J. A. Perri, S. Laplaca, B. Post, *Acta Cryst.* **11**, 310 (1958).
20. S. Yue *et al.*, *Mater. Today Phys.* **13**, 100194 (2020).
21. A. Rai, S. Li, H. L. Wu, B. Lv, D. G. Cahill, *Phys. Rev. Mater.* **5**, 013603 (2021).
22. A. A. Maznev, T. F. Crimmins, K. A. Nelson, *Opt. Lett.* **23**, 1378–1380 (1998).
23. A. A. Maznev, K. A. Nelson, J. A. Rogers, *Opt. Lett.* **23**, 1319–1321 (1998).
24. S. Huberman *et al.*, *Science* **364**, 375–379 (2019).
25. K. Chen *et al.*, *Carbon* **107**, 233–239 (2016).
26. J. L. Lyons *et al.*, *Appl. Phys. Lett.* **113**, 251902 (2018).
27. M. Fava *et al.*, *Npj Comput. Mater.* **7**, 54 (2021).
28. M. Fava *et al.*, How dopants limit the ultrahigh thermal conductivity of boron arsenide: A first principles study, version 1, Zenodo (2021); https://doi.org/10.5281/zenodo.4453192.
29. C. Freysoldt *et al.*, *Rev. Mod. Phys.* **86**, 253–305 (2014).
30. F. Tian *et al.*, *Appl. Phys. Lett.* **114**, 131903 (2019).
31. X. Chen *et al.*, *Phys. Rev. Appl.* **11**, 064070 (2019).
32. N. D. Arora, J. R. Hauser, D. J. Roulston, *IEEE Trans. Electron Dev.* **29**, 292–295 (1982).

ACKNOWLEDGMENTS

Funding: This work was supported by the Office of Naval Research under Multidisciplinary University Research Initiative grant N00014-16-1-2436. This work made use of the MRSEC Shared Experimental Facilities at MIT, supported by the National Science Foundation under award no. DMR-1419807. **Author contributions:** Conceptualization: J.S., G.A.G., K.C., Z.R., and G.C. Materials preparation: G.A.G. and F.T. TG experiments: J.S. and K.C. TG modeling: J.S., K.C., X.Q., and G.C. Time-domain thermoreflectance, frequency-domain thermoreflectance, PL, and Raman measurements: J.S. Hall measurements: H.L., Jian.Z., L.S., T.N., F.H., and M.L. Mobility calculations: Z.D. and Jiaw.Z. Data analysis: J.S., G.A.G., Z.D., K.C., D.B., Z.R., and G.C. Funding and project administration: Z.R. and G.C. Supervision: Z.R. and G.C. Writing – original draft: J.S., G.A.G., Z.D., Z.R., and G.C. Review & editing: all authors. **Competing interests:** US patent publication no. 20210269318 has been granted to the method for growing c-BAs crystals presented in this work. A patent application on the high mobility of Bas, entitled "Ultra-high ambipolar mobility in cubic arsenide," has been filed with the US Patent Office. The authors declare no other competing interests. **Data and materials availability:** All data are available in the main text and supplementary materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn4290
Materials and Methods
Supplementary Text
Figs. S1 to S12
Table S1
References (*33*–*55*)

Submitted 1 December 2021; accepted 16 June 2022
10.1126/science.abn4290

By **Sophia X. Pfister**

From weakness comes strength

It was 3 a.m. I was exhausted from taking care of my 3-month-old baby, but I couldn't sleep. As I tried to recall the topics of the five conference calls on my calendar for the morning, I again had the haunting thought that I wasn't good enough for my job—a director position I started shortly before my baby was born. I imagined I would make mistakes in my presentations and my team would lose respect for me. Tormented by these thoughts, I reached for a book from the pile on my bedside table to distract myself. By chance I grabbed the Bible, which I had been too busy to read since my baby was born. As I opened it to a random page and happened on the verse "For when I am weak, then I am strong," tears filled my eyes, and I could breathe again.

My upbringing gave me an "achiever" personality. From childhood class president to prestigious university degrees to a leadership position in a large company, I was regarded as a "star." People see me as confident, ambitious, competent, and energetic. But I always feared seeming imperfect in the eyes of others. I worked as hard as I could to make up for my flaws.

But after becoming a new mom and starting a new job, I was unable to excel no matter how hard I worked. The job required me to attend meetings with almost no break between 7 a.m. and 5 p.m., pushing my own work tasks late into the night. I used a breast pump under the table during meetings and frequently forgot to eat. Mental and physical exhaustion from back-to-back meetings and lack of sleep made it difficult to think deeply and creatively about science. I wanted to offer useful comments in meetings, but my thoughts often became muddled, at times leaving me tongue-tied midsentence. I became so anxious about my long to-do list that I could not calm down to tackle a single task. When a team member left for a new job, I blamed myself. On top of it all, I developed postpartum depression but was too ashamed to tell my doctor. It was the lowest point of my life, and I could no longer deny my weakness.

After the 3 a.m. epiphany, I wrote the Bible verse on a sticky note and put it on the corner of my computer as a reminder. I read it to myself as I transitioned from one meeting to the next, and it began to transform my approach to work. I realized that instead of focusing on trying to make "clever" comments in meetings—and feeling stressed that I couldn't come up with any—I could acknowledge what I didn't know and ask honest questions to learn from others. And when I got overwhelmed by my long to-do list, I learned to accept my limits, identify the most important tasks, and trust my team by delegating.

"I shared my struggles as a new mom and ... my team members opened up."

Accepting my weakness also helped me find a path to more authentic leadership. I previously put my own and others' feelings in a box, thinking that discussing them would distract from our productivity, and instead focused on data, timelines, and deliverables. But after my own crisis, I began to pay more attention to my team members' emotional well-being. I shared my struggles as a new mom and my fear of not finding the best direction for the team. In response, my team members opened up to me about the challenges they were facing. These conversations helped build trust, loyalty, and team morale.

I also became less judgmental when I had to give critical feedback to team members. Previously, I saw it as a persuasion contest to convince them to stop doing things their way and adopt the "right way," and I dreaded doing it. But now, I first seek to understand the motivation behind their behavior. This enables me to deliver feedback with the aim of helping each individual become their best self.

Now, I am grateful for my weaknesses, as they make me humbler. They taught me that true strength, in life and in leadership, does not rely on authority and power, but on compassion, honesty, and kindness. ■

Sophia X. Pfister is the director of research science at Varian, a Siemens Healthineers company in Palo Alto, California. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

CALL FOR PAPERS

SPJ
SCIENCE PARTNER JOURNALS

北京理工大學
BEIJING INSTITUTE OF TECHNOLOGY

中国空间技术研究院
China Academy of Space Technology

Space: Science & Technology

Space: Science & Technology is an online-only, Open Access journal published in affiliation with **Beijing Institute of Technology (BIT)** and distributed by the **American Association for the Advancement of Science (AAAS)**. BIT cooperates with China Academy of Space Technology (CAST) in managing the journal. The journal promotes the exploration and research of space worldwide, to lead the rapid integration and technological breakthroughs of interdisciplinary sciences in the space field, and to build a high-level academic platform for discussion, cooperation, technological progress and information dissemination among professional researchers, engineers, scientists and scholars.

Submit your research to *Space: Science & Technology* today!

Learn more at **spj.sciencemag.org/space**

The Science Partner Journal (SPJ) program was established by the American Association for the Advancement of Science (AAAS), the nonprofit publisher of the *Science* family of journals. The SPJ program features high-quality, online-only, Open Access publications produced in collaboration with international research institutions, foundations, funders and societies. Through these collaborations, AAAS furthers its mission to communicate science broadly and for the benefit of all people by providing top-tier international research organizations with the technology, visibility, and publishing expertise that AAAS is uniquely positioned to offer as the world's largest general science membership society.

Visit us at **spj.sciencemag.org**

@SpaceSciTech @SPJournals

OPEN ACCESS

ARTICLE PROCESSING CHARGES WAIVED UNTIL JULY 2023