Can a fabled blue parrot flourish again in the wild? p. 1148

Unlocking hidden physics with quantum computers pp. 1154 & 1182

Inherited microbiomes are shaped by lifestyle p. 1220

Science

$15
10 JUNE 2022
SPECIAL ISSUE
science.org
AAAS

NUCLEAR PORE

A gateway complex

p. 1172



62% Portion of 1640 clinical trials rated as "bad," defined as those at high risk of bias because of selective reporting of results and other flaws. The bad ones wasted as much as £8 billion. (*Trials*)

IN BRIEF

Edited by **Jeffrey Brainard**



Pigs are housed at a farm on the outskirts of Hanoi as Vietnam works to counter swine disease.

AGRICULTURE

Vaccine targets African swine fever

Vietnam's agriculture ministry last week gave limited authorization to a vaccine hailed as an important tool to control one of the most serious animal diseases, African swine fever (ASF). In recent years the sickness has hit pig herds hard in several Asian and European countries. Vietnam's National Veterinary Joint Stock Company developed the vaccine based on an ASF virus strain engineered by the U.S. Agricultural Research Service to lack a gene linked to virulence. A small trial of 20 animals, reported in September 2021, found strong evidence of protection; the company says an unpublished, follow-up trial of 131 pigs showed 99% of those that got full doses survived ASF infections. Based on these results, the ministry approved commercial use of the vaccine in up to 600,000 pigs. It will evaluate results before deciding whether to allow nationwide use. Endemic in Africa, ASF spread through much of Europe in the 2000s and to Asia in 2018, requiring culling and creating shortages of pork, a major source of protein throughout the region.

FDA panel backs Novavax vaccine

COVID-19 | The U.S. Food and Drug Administration's (FDA's) vaccine advisory panel this week recommended nearly unanimously that the agency authorize a protein-based COVID-19 vaccine from Novavax, which would be the first of its kind available to U.S. adults. Panel members said benefits of the vaccine, made of the SARS-CoV-2 spike protein combined with an immune-boosting substance, outweighed risks when it is given in two doses 3 weeks apart to those 18 years and older. FDA doesn't have to abide by its advisers' recommendations but usually does. In a 30,000-person trial in the United States and Mexico, the vaccine was 90.4% efficacious at preventing symptomatic infection by early strains of SARS-CoV-2. The approval came days after FDA posted data documenting five cases of myocarditis or pericarditis—inflammations of heart tissue—in volunteers, most of them young men, soon after they received the vaccine in U.S. and U.K. clinical trials. Novavax hopes its product will attract U.S. recipients skeptical of vaccines that employ messenger RNA and booster seekers who favor its proven method, which has led to licensed vaccines for other diseases, such as shingles.

Studies of low-dose radiation urged

BIOMEDICINE | The U.S. government should spend $100 million per year for at least 15 years to study the health effects of low-dose radiation, a high-profile review panel concluded last week. The public and workers are routinely exposed to low-dose radiation (below 100 milligray, a measurement of absorbed dose) from sources such as medical scans, air travel, and mining, which contributes to cancer and possibly heart disease and other health problems. The Department of Energy's (DOE's) Office of Science ended a long-running program to study low-dose radiation in 2016 so it could focus on other priorities. But in 2018, Congress mandated its revival and later asked the National Academies of Sciences, Engineering, and Medicine for a new blueprint. The research is important and should resume, although not entirely

PHOTO: NHAC NGUYEN/AFP/GETTY IMAGES

under DOE's sponsorship, the academies' report says, noting the conflicts of interest associated with its nuclear weapons facilities. The report recommends the National Institutes of Health fund epidemiological and biological studies; DOE should oversee computational work and modeling. Congress must now decide whether to appropriate the funding.

Bees gets protected as 'fish'

CONSERVATION | Four species of bumble bees qualify for protection under California's Endangered Species Act because they fit a loophole in the state's definition of "fish," an appeals court ruled last week. Until now, the state law protected no insect species. But a state Court of Appeal based in Sacramento pointed to California's Fish and Game Code, which includes in the definition of fish any "mollusk, crustacean, invertebrate, (or) amphibian." That wording covers any terrestrial invertebrate, such as a bumble bee, the court wrote. The ruling was celebrated by conservation organizations and bemoaned by agricultural groups, which argued that extending the protection to the bumble bees would burden farming operations. Bee populations have declined across the United States and elsewhere, posing threats to agricultural crops and other plants that depend on pollinators for healthy development.

NIH grantees lax on foreign detail

RESEARCH SECURITY | A U.S. government watchdog has found that many institutions receiving funding from the National Institutes of Health (NIH) don't follow federal rules on reporting foreign sources of support, educating scientists about those rules, and investigating possible conflicts of interest. A 22 June report by the inspector general of NIH's parent department found, for example, that 36% of the more than 600 institutions surveyed in late 2020 don't require their faculty members to disclose participation in another country's talent recruitment program and 37% don't distinguish between domestic and foreign funding. Since 2018, NIH has been especially vigilant in tracking grantees' links to China as part of a governmentwide campaign to prevent the theft of U.S.-funded research by that country. The report calls on NIH to enforce the existing rules, which institutions must obey as a condition of funding.

Rice led to chicken domestication

EVOLUTION | People around the world know chicken and rice is a winning culinary combination. But now, scientists say without rice, there might not have been chickens. It wasn't until humans began clearing forest and sowing rice seeds within the range of red jungle fowl in Southeast Asia that some of these wild pheasants swept down from the trees to feed on the seeds—and evolved into more docile chickens, according to research published this week in the *Proceedings of the National Academy of Sciences*. This taming of the jungle fowl happened much more recently than other studies have estimated, according to the comprehensive analysis of bones and dates at more than 600 sites, which found that some bones thought to be chickens belonged to other animals. The authors say the oldest chickens appear just 3250 to 3650 years ago at a rice farming site in what is now central Thailand. Then, chickens spread across Asia with rice and millet farming.



IN FOCUS Egypt's antiquities ministry last week unveiled a new collection of artifacts from its Late Period (about 664 B.C.E. to 332 B.C.E.) found within the Saqqara necropolis, near Cairo. The new discoveries from the previously excavated cemetery include 150 bronze statues of ancient Egyptian deities and 250 wooden sarcophagi.

PHOTO: MAHMOUD EL-KHAWAS/PICTURE ALLIANCE/GETTY IMAGES

IN DEPTH



Canada has begun to offer a monkeypox vaccine made by Bavarian Nordic to select groups, including contacts of known cases of the disease.

PHOTO: CHRISTINNE MUSCHI/REUTERS

INFECTIOUS DISEASE

Monkeypox vaccination plans take shape amid questions

Favored shot is a seemingly safer smallpox vaccine, but researchers debate how best to use it

By **Kai Kupferschmidt**

In 1959, German microbiologist Anton Mayr took a strain of vaccinia, a poxvirus used to inoculate against smallpox, and started to grow it in cells taken from chicken embryos. After several years of transferring the strain to fresh cells every few days, the virus had changed so much it could no longer reproduce in most cells from mammals. But it could still produce an immune response that protected against smallpox.

Mayr had set out to study how poxviruses evolve, but by accident he had produced a potentially safer smallpox vaccine. Dubbed Modified Vaccinia Ankara (MVA) because the original viral strain came from that Turkish city, the vaccine had a short career. "With smallpox eradicated in 1980, it disappeared into the freezer," says Gerd Sutter, a virologist at the Ludwig Maximilian University of Munich, who has studied Mayr's vaccinia strain for decades.

Now, this virus, further weakened and brought to the market by the Danish pharma company Bavarian Nordic, may become key to arresting the largest outbreak of monkeypox ever seen outside Africa, which has already sickened more than 1000 people. It is the only vaccine licensed anywhere for use against monkeypox, although other, riskier smallpox vaccines also appear to work. The United States, the United Kingdom, Canada, and several other countries have already started to "ring" vaccinate, offering it to contacts of identified monkeypox cases, including health care workers and sexual partners. "MVA will be very important in this outbreak because it is a nonreplicating vaccine, which means it doesn't have the same side effect profile as some of the other live [virus] vaccines" being considered, says Rosamund Lewis, technical lead on monkeypox at the World Health Organization (WHO).

But what role the vaccine will ultimately play depends on a host of factors: whether those most at risk from infection can be identified and vaccinated, whether the vaccine is as effective as hoped, and whether enough is available to stop the burgeoning outbreak. WHO has so far only backed ring vaccination—MVA is ideally given within 4 days of an exposure but recommended for up to 14 days—but some scientists say it's too difficult to reach the specific contacts people had. They advocate broader vaccination campaigns in the population most affected so far: men who have sex with men (MSM).

Hundreds of millions of doses of smallpox vaccine are stored around the world, insurance against a possible release of the dreaded virus by terrorists or in war, and they are known to offer some protection against monkeypox. A study in the Democratic Republic of the Congo (DRC) in the 1980s found that household contacts of people sick with monkeypox were seven times less likely to contract the disease if they had been vaccinated against smallpox. Yet the vast majority of existing smallpox vaccines consist still replicating vaccinia. These can cause rare but life-threatening side effects such as a encephalitis or progressive vaccinia, the spread of the vaccine virus to the whole body, to which immunocompromised people are vulnerable.

Although 66 people have already died of monkeypox this year in African countries, the recent cases in nonendemic countries have mostly been mild. And many contacts of those infected are living with HIV, which could make them more likely to suffer from vaccinia side effects. Given the risks and benefits, "using these vaccines is out of the question," Sutter says.

Bavarian Nordic's nonreplicating vaccine, marketed as Jynneos in the United States and as Imvanex in Europe, sidesteps some of the risk. So does a vaccinia-based vaccine named LC16m8, licensed for smallpox only in Japan, which also appears to cause fewer side effects. "I believe these

are the ones that are going to be used [in the new outbreak] because they have a much-enhanced safety profile," says Marion Gruber, who headed the U.S. Food and Drug Administration's vaccine office until October 2021.

Canada and the United States have already licensed MVA for use against monkeypox and Bavarian Nordic is in talks with the European Medicines Agency (EMA). "I really hope that in a matter of 1 or 2 months from now, this can be approved" in Europe, says EMA's Marco Cavaleri.

The United Kingdom has been using MVA "off-label" for a few years to vaccinate contacts of imported monkeypox cases. WHO's Strategic Advisory Group of Experts on Immunization is set to release guidance in the next days that will back MVA, but it will also recommend using earlier vaccines in certain scenarios. Still, Cavaleri says, "If [MVA] is available, clearly that will be the vaccine to start."

Exactly how much is available remains murky. "Countries have been reluctant over the past couple of decades to share that information in detail with WHO but WHO is now reaching out to all of them again," Lewis says. The United States, which supported development of MVA, likely has the biggest supply. A federal spokesperson says the Strategic National Stockpile has 36,000 doses, that another 36,000 doses will be delivered "in the near future," and that the company is storing bulk material for millions more U.S.-reserved doses. A Bavarian Nordic spokesperson says many other countries had ordered its MVA vaccine in the past weeks and the company was trying to send smaller batches to countries "so they can start to vaccinate sooner rather than later."

How broadly to roll out MVA, or any vaccine, remains the key debate. Ring vaccination among MSM can be challenging given the stigma faced by that group in many cultures and the nature of the contacts. A paper published last week in *Eurosurveillance* noted that in the United Kingdom, many of those infected reported sexual contacts with people whose details they either did not know or did not want to share.

The Canadian province of Quebec has already extended vaccinations from direct contacts of monkeypox cases to any men who have had more than two male sexual partners in the past 14 days. Another way to tackle the problem, says Yale School of Public Health epidemiologist Gregg Gonsalves, "would be to offer it to individuals who have attended social events in which close contact with someone infected was possible, but this would increase the numbers of those being recruited for vaccination even further." Germany is also likely to offer the vaccine more broadly, though it will not quickly have enough vaccine available for all MSM, says Leif Erik Sander, an infectious disease expert at the Charité University Hospital in Berlin.

Even where contacts of infected cases were identified, uptake has been low. The same U.K. study reported that 169 out of 245 health care workers who had been offered MVA had taken it, but only 15 out of 107 contacts in other groups. "It's very challenging to target high-risk groups while balancing stigma and encouraging uptake of the vaccines," says Boghuma Titanji, a virologist at Emory University. The politicization of vaccines during COVID-19 has increased the barriers, she adds.

> **"The truth is, we don't know the efficacy of any of these monkeypox vaccines."**
>
> **Ira Longini**, University of Florida

How well MVA really protects humans from monkeypox is uncertain. The license for MVA in Canada and the United States is based on animal studies, where it was shown to protect macaques and prairie dogs, plus data in humans showing a strong antibody response. A pair of DRC studies vaccinated 1600 health care workers with one of two MVA formulations and found no monkeypox cases in each 2-year study period. But there were no control groups, and one vaccinated health care worker did get monkeypox half a year later. "The truth is, we don't know the efficacy of any of these monkeypox vaccines," says Ira Longini, a biostatistician at the University of Florida who is advising WHO.

That is why WHO has urged countries that deploy monkeypox vaccine to study how well it works and how best to use it. "If we want to contain these outbreaks and learn something about the efficacy of these vaccines, it's going to have to be a concerted effort with protocols and organized properly," Longini says. One question is whether a single dose of the vaccine, which is normally given as two doses 4 weeks apart, may suffice. That could encourage more uptake and stretch supplies.

The question of vaccine equity looms large, too. Titanji notes that the hopes for MVA are based partly on the DRC data. "It's almost a moral obligation to make sure that, if these vaccines are being utilized elsewhere now, the people on whom the data was generated, who have been dealing with monkeypox for 50 years, should have access to it as well." ■

RESEARCH ETHICS

Chile's Indigenous groups seek fairer research

New constitution may help reset relationship between scientists and communities

By **Emiliano Rodríguez Mega**

The small fishing settlement of Puerto Edén is nestled on Wellington Island in southern Chile, among a labyrinth of islets and fjords at least a day's journey from the nearest city. But the distance and Patagonian cold have not discouraged generations of scientists from making the trip. Puerto Edén is home to some of the Kawésqar, descendants of nomadic seafarers. Their culture, territory, the remains of their ancestors, and their dying language have all drawn academic interest.

But the goals of researchers and the community have sometimes been at odds, says Ayelen Tonko Huenucoy, a Kawésqar physical anthropologist at the Chilean National Museum of Natural History, who partly grew up in Puerto Edén. "Several scientists arrived in a totally conquerorlike way … using us for their [own] goals," such as demanding genetic information from the community, she says.

Now, the Kawésqar and other Indigenous peoples in Chile hope to see their rights recognized for the first time in the country's new constitution, which Chileans will vote on in a referendum in September. And other efforts to balance the relationship between Indigenous groups and scientists in Chile are underway, including a collaborative workshop last week on ethics and genomics.

"We will no longer be the guinea pigs," says Elisa Loncón, a Mapuche linguist at the University of Santiago and former president of the constitutional convention. "And we will not be a hindrance to knowledge either."

The constitutional process began in 2019, when massive protests against inequality called for replacing the constitution enacted during Augusto Pinochet's dictatorship in 1980. If approved, the new constitution would make Chile "plurinational," with at least 11 Indigenous groups, representing

are the ones that are going to be used [in the new outbreak] because they have a much-enhanced safety profile," says Marion Gruber, who headed the U.S. Food and Drug Administration's vaccine office until October 2021.

Canada and the United States have already licensed MVA for use against monkeypox and Bavarian Nordic is in talks with the European Medicines Agency (EMA). "I really hope that in a matter of 1 or 2 months from now, this can be approved" in Europe, says EMA's Marco Cavaleri.

The United Kingdom has been using MVA "off-label" for a few years to vaccinate contacts of imported monkeypox cases. WHO's Strategic Advisory Group of Experts on Immunization is set to release guidance in the next days that will back MVA, but it will also recommend using earlier vaccines in certain scenarios. Still, Cavaleri says, "If [MVA] is available, clearly that will be the vaccine to start."

Exactly how much is available remains murky. "Countries have been reluctant over the past couple of decades to share that information in detail with WHO but WHO is now reaching out to all of them again," Lewis says. The United States, which supported development of MVA, likely has the biggest supply. A federal spokesperson says the Strategic National Stockpile has 36,000 doses, that another 36,000 doses will be delivered "in the near future," and that the company is storing bulk material for millions more U.S.-reserved doses. A Bavarian Nordic spokesperson says many other countries had ordered its MVA vaccine in the past weeks and the company was trying to send smaller batches to countries "so they can start to vaccinate sooner rather than later."

How broadly to roll out MVA, or any vaccine, remains the key debate. Ring vaccination among MSM can be challenging given the stigma faced by that group in many cultures and the nature of the contacts. A paper published last week in *Eurosurveillance* noted that in the United Kingdom, many of those infected reported sexual contacts with people whose details they either did not know or did not want to share.

The Canadian province of Quebec has already extended vaccinations from direct contacts of monkeypox cases to any men who have had more than two male sexual partners in the past 14 days. Another way to tackle the problem, says Yale School of Public Health epidemiologist Gregg Gonsalves, "would be to offer it to individuals who have attended social events in which close contact with someone infected was possible, but this would increase the numbers of those being recruited for vaccination even further." Germany is also likely to offer the vaccine more broadly, though it will not quickly have enough vaccine available for all MSM, says Leif Erik Sander, an infectious disease expert at the Charité University Hospital in Berlin.

Even where contacts of infected cases were identified, uptake has been low. The same U.K. study reported that 169 out of 245 health care workers who had been offered MVA had taken it, but only 15 out of 107 contacts in other groups. "It's very challenging to target high-risk groups while balancing stigma and encouraging uptake of the vaccines," says Boghuma Titanji, a virologist at Emory University. The politicization of vaccines during COVID-19 has increased the barriers, she adds.

> **"The truth is, we don't know the efficacy of any of these monkeypox vaccines."**
>
> **Ira Longini**, University of Florida

How well MVA really protects humans from monkeypox is uncertain. The license for MVA in Canada and the United States is based on animal studies, where it was shown to protect macaques and prairie dogs, plus data in humans showing a strong antibody response. A pair of DRC studies vaccinated 1600 health care workers with one of two MVA formulations and found no monkeypox cases in each 2-year study period. But there were no control groups, and one vaccinated health care worker did get monkeypox half a year later. "The truth is, we don't know the efficacy of any of these monkeypox vaccines," says Ira Longini, a biostatistician at the University of Florida who is advising WHO.

That is why WHO has urged countries that deploy monkeypox vaccine to study how well it works and how best to use it. "If we want to contain these outbreaks and learn something about the efficacy of these vaccines, it's going to have to be a concerted effort with protocols and organized properly," Longini says. One question is whether a single dose of the vaccine, which is normally given as two doses 4 weeks apart, may suffice. That could encourage more uptake and stretch supplies.

The question of vaccine equity looms large, too. Titanji notes that the hopes for MVA are based partly on the DRC data. "It's almost a moral obligation to make sure that, if these vaccines are being utilized elsewhere now, the people on whom the data was generated, who have been dealing with monkeypox for 50 years, should have access to it as well." ■

RESEARCH ETHICS

Chile's Indigenous groups seek fairer research

New constitution may help reset relationship between scientists and communities

By **Emiliano Rodríguez Mega**

The small fishing settlement of Puerto Edén is nestled on Wellington Island in southern Chile, among a labyrinth of islets and fjords at least a day's journey from the nearest city. But the distance and Patagonian cold have not discouraged generations of scientists from making the trip. Puerto Edén is home to some of the Kawésqar, descendants of nomadic seafarers. Their culture, territory, the remains of their ancestors, and their dying language have all drawn academic interest.

But the goals of researchers and the community have sometimes been at odds, says Ayelen Tonko Huenucoy, a Kawésqar physical anthropologist at the Chilean National Museum of Natural History, who partly grew up in Puerto Edén. "Several scientists arrived in a totally conquerorlike way … using us for their [own] goals," such as demanding genetic information from the community, she says.

Now, the Kawésqar and other Indigenous peoples in Chile hope to see their rights recognized for the first time in the country's new constitution, which Chileans will vote on in a referendum in September. And other efforts to balance the relationship between Indigenous groups and scientists in Chile are underway, including a collaborative workshop last week on ethics and genomics.

"We will no longer be the guinea pigs," says Elisa Loncón, a Mapuche linguist at the University of Santiago and former president of the constitutional convention. "And we will not be a hindrance to knowledge either."

The constitutional process began in 2019, when massive protests against inequality called for replacing the constitution enacted during Augusto Pinochet's dictatorship in 1980. If approved, the new constitution would make Chile "plurinational," with at least 11 Indigenous groups, representing



Linguist Elisa Loncón carries the Mapuche flag at Chile's constitutional convention in 2021 in Santiago.

more than 2 million people or nearly 13% of the population, recognized as autonomous communities governing their territories. They would in theory have more sway over their lands, than, for example, Native Americans in the United States, where the federal government holds Indigenous land in trust.

The draft constitution recognizes the existence of Indigenous knowledge and protects Indigenous peoples' identities, cultures, and territories, including nature in its "material and immaterial dimensions." It also gives Indigenous peoples the right to repatriate objects and human remains, and mandates that the Chilean government develop mechanisms for such repatriation, perhaps including objects from abroad.

The new constitution isn't explicit about research with Indigenous communities. But it could encourage a more collaborative approach that considers local and ancestral knowledge, says microbiologist Cristina Dorador Ortiz, a member of the constitutional convention that wrote it.

This stance is new in Chile, where some Indigenous people cite past examples of scientific overreach. "Many times, communities complain that research is done on them from a Western perspective," Dorador Ortiz says. For example, in the 1990s, Chilean and Japanese researchers took blood from Huilliche communities, who are part of the Mapuche people, in southern Chile. Those samples and more than 3500 others from Indigenous groups across South America are now in a public cell bank at the RIKEN BioResource Research Center in Tsukuba, Japan. Cell lines derived from the samples, expected to be useful for studies on human migration or genetic variations in drug response, are available to scientists worldwide, with a tube costing about $110. But donors never saw any benefits, Tonko Huenucoy says.

It's a story familiar to others in Chile. "The way research is done nowadays is superconvenient" for scientists, says Constanza Silva Gallardo, a biological anthropologist at Pennsylvania State University, University Park, and a member of the Diaguita Mapochogasta Autonomous Community in Santiago. "There needs to be some sort of pushback to bring effective change."

The proposed constitution could help set the stage, although polls suggest its initial high popularity has recently fallen. But even if it fails, other efforts are ongoing. In March, a mostly Chilean team including Tonko Huenucoy and Silva Gallardo published a paper in *Frontiers in Genetics* urging geneticists to abandon stigmatizing narratives that magnify any genetic differences between Indigenous people and other Chileans. They also called for Chilean universities to develop protocols to incorporate Indigenous voices in designing sampling procedures, drafting informed consent forms, and interpreting results.

In late 2021, this same group launched a program, Ciencia y Comunidades, to improve ethical standards in genomic studies of Indigenous populations in Chile. Last week, they held a workshop at the Pontifical Catholic University of Chile with members of the Aymara, Diaguita, Colla, Chango, Rapa Nui, and Mapuche (including Huilliche and Pehuenche) peoples. Between opening and closing ceremonies involving traditional dances, attendees discussed how research is done, who approves projects, and what genetic data can and cannot say about a person's identity. The effort was modeled after the Summer Internship for Indigenous Peoples in Genomics workshop, an international consortium that explores the ethics of genomics and aims to train Indigenous scientists in the field (*Science*, 28 September 2018, p. 1304).

Cape Town meeting slams 'helicopter research'

By **Cathleen O'Grady**

When researchers from wealthy countries engage in "helicopter research"—field research in poorer countries that extracts data without respectful collaboration—they violate research integrity as well as pose a moral problem, said attendees at last week's World Conference on Research Integrity, held in Cape Town, South Africa.

The conference saw the launch of the Cape Town Statement on equitable research partnerships, which attendees will finalize and submit to an academic journal. Those at the meeting hope their new framing will elevate the issue and help spur systemic solutions, rather than leaving the task of building fair collaborations up to individual researchers.

Researchers in low- and middle-income countries (LMICs) often feel unappreciated when they partner with researchers from wealthier countries, Francis Kombe, co-chair of the African Research Integrity Network and a contributor to the statement, told the conference. Local experts are too often not listed as authors, cannot access data they gathered, and lack the power to steer research to local priorities, studies of the issue have found. All this can affect the quality of research.

Such "scientific colonialism" uses the same tactics as colonialism has historically, Sue Harrison, deputy vice-chancellor for research and internationalization at the University of Cape Town, said at the event. It extracts data instead of raw materials—and undermines and underfunds local infrastructure and skills. This leaves researchers in LMICs without the publications, patents, and skills of their wealthier counterparts.

The Cape Town Statement will offer a guide for how institutions can improve collaborations. For example, funders could set out expectations for equal authorship and data access, says Minal Pathak, a climate researcher at Ahmedabad University in India.

She hopes the statement has impact. "Maybe it's not new. But maybe we need to say it another time."

PHOTO: JAVIER TORRES/AFP/GETTY IMAGES

The goal is to empower communities to demand their rights and to "motivate colleagues to work in a different way," says Constanza de la Fuente, a Chilean ancient DNA researcher at the University of Chicago and a member of Ciencia y Comunidades. "To approach the communities not only saying, 'This is an informed consent form, sign it and give me your sample,' but trying to generate a dialogue with them."

Although they acknowledge that dialogue is needed, some Chilean researchers are wary. In other countries, including Canada, New Zealand, and the United States, Indigenous communities have asked geneticists to delay work, change research questions, keep data private, and not publish results. "One can't take absolutist [attitudes]," such as insisting that scientists must keep all data private, says Lucía Cifuentes, a medical geneticist at the University of Chile (UCh), Santiago. "Science needs creative freedom."

Publishing restrictions would be "censorship," says Ricardo Verdugo, a human population geneticist also at UCh Santiago. But he thinks a new paradigm is needed. Indigenous communities are "the first ones that have the right to have a voice," he says. "What to ask, why ask it, and how I'm going to interpret [and] communicate it, is something that absolutely requires [their] opinion."

For others, now is the moment for drastic measures. "Other scientists might [question] me. But, for me, ethics comes first," says Macarena Fuentes, a human population geneticist at the University of Tarapacá, Arica. "For there to be a transition, extreme changes must occur."

In Puerto Edén, fed up with what they saw as one-sided interactions, the community created a protocol for scientific research within its territory. Scientists must meet with a council to explain their research, what they'll do with the results, and how Puerto Edén will benefit. They must also respect Kawésqar culture, including honoring taboos against visiting sacred places. And they must give something back, whether a simple acknowledgement, a share in any financial rewards, or co-authorship. The plan may be exceptional in Chile at the moment, but many hope it will become the norm in the future.

The protocol isn't a rejection of science, Tonko Huenucoy explains. The community even plans to build a science center and field station to attract research to the community. But they want to make sure that it's done for and with the Kawésqar, she says. So "[our] voices are included from the very beginning." ■

Emiliano Rodríguez Mega is a journalist in Mexico City.

RESEARCH INTEGRITY

Controversial botanist cleared

Report sees "insufficient evidence" of misconduct

By **Charles Piller**

When eight scientists filed a misconduct complaint against prominent botanist Steven Newmaster with the University of Guelph (UG) in June 2021, they thought they had an ironclad case. Their claim that Newmaster, whose work profoundly influenced how dietary supplements are tested and marketed, had made up or plagiarized data in three papers was "an entirely credible and well-founded allegation," says cosignatory Kenneth Thompson, a postdoc at Stanford University.

But an investigative committee at UG disagrees. Newmaster "displayed a pattern of poor judgement," his conduct was "suspicious," and there were "many shortcomings" in his work, panel chairman John Walsh, a business professor at UG, wrote in a 1 June letter to the complainants. But there was "insufficient evidence" to find Newmaster guilty of misconduct.

"Given the evidence of data falsification that was assembled, I was very surprised at the conclusion," says evolutionary biologist Paul Hebert, a co-signatory to the complaint. Hebert directs UG's Centre for Biodiversity Genomics and pioneered DNA barcoding, a technique for identifying organisms from short snippets of DNA that is central to Newmaster's work.

> **"I was very surprised at the conclusion."**
> **Paul Hebert**, University of Guelph

But Thomas Braukmann, a former postdoc at the UG center who is now at Public Health Ontario, says the findings are "disappointing but not surprising." "I don't think [UG] took the allegations seriously or put the committee together in good faith," says Braukmann, who was not involved in the case but studied Newmaster's papers at *Science*'s request. "We need a better system in Canada to handle misconduct concerns." Walsh and UG did not respond to requests for comment.

Newmaster made headlines with a 2013 *BMC Medicine* study reporting that many herbal supplements didn't contain the labeled ingredients and some had toxic contaminants. The paper propelled him to global fame as a testing expert. His own companies and a nonprofit group at UG raised millions of dollars by certifying supplements, cannabis, and other comestibles.

Thompson was the first to level accusations against Newmaster, in 2020. He claimed Newmaster made up and plagiarized the data for a study testing the ability of DNA barcoding to identify species in a Canadian forest, which they published together in 2014, when Thompson was a UG undergraduate. After the university dismissed his complaint, the group of eight sent a more elaborate complaint letter. It fingered not only Thompson's paper—which the journal *Biodiversity and Conservation* retracted in October 2021—but also the supplements study and a 2013 paper using DNA barcoding to study woodland caribou diets. Newmaster's co-authors have requested retractions of those articles as well.

For now, *BMC Medicine* is investigating the allegations about the supplements article and the Canadian *Journal of Forest Research* has added an Expression of Concern to the caribou paper. Newmaster has denied the charges. "I have never engaged in any unethical activity or academic misconduct," he wrote in an official reply obtained by *Science*.

An investigation by *Science* (3 February, p. 484) revealed many other cases in which Newmaster appeared to manipulate or fabricate data, plagiarize, and invent elements of his academic record. He did not respond to requests for comment for the *Science* story, and the panel did not address the issues it raised. UG could take up to several months to issue a final decision.

Newmaster's accusers had been worried the investigation might not be rigorous, given the panel's lack of expertise in genomics. In addition to Walsh, it included Jeff Wichtel, dean of UG's veterinary college, and Cynthia Fekken, a psychologist from Queen's University. Walsh's letter says they relied on an independent expert witness, whom he did not name. The letter notes that a "key factor" in the panel's finding that it was impossible to "definitely establish" misconduct was the "absence of records, including raw data."

That's a particularly frustrating argument, Thompson says. "That was the essence of our complaint. We knew they wouldn't be able to find records. Our complaint alleged that Prof. Newmaster falsified his work and never had the data to back it up." ■

This story was supported by the *Science* Fund for Investigative Reporting.



Linguist Elisa Loncón carries the Mapuche flag at Chile's constitutional convention in 2021 in Santiago.

more than 2 million people or nearly 13% of the population, recognized as autonomous communities governing their territories. They would in theory have more sway over their lands, than, for example, Native Americans in the United States, where the federal government holds Indigenous land in trust.

The draft constitution recognizes the existence of Indigenous knowledge and protects Indigenous peoples' identities, cultures, and territories, including nature in its "material and immaterial dimensions." It also gives Indigenous peoples the right to repatriate objects and human remains, and mandates that the Chilean government develop mechanisms for such repatriation, perhaps including objects from abroad.

The new constitution isn't explicit about research with Indigenous communities. But it could encourage a more collaborative approach that considers local and ancestral knowledge, says microbiologist Cristina Dorador Ortiz, a member of the constitutional convention that wrote it.

This stance is new in Chile, where some Indigenous people cite past examples of scientific overreach. "Many times, communities complain that research is done on them from a Western perspective," Dorador Ortiz says. For example, in the 1990s, Chilean and Japanese researchers took blood from Huilliche communities, who are part of the Mapuche people, in southern Chile. Those samples and more than 3500 others from Indigenous groups across South America are now in a public cell bank at the RIKEN BioResource Research Center in Tsukuba, Japan. Cell lines derived from the samples, expected to be useful for studies on human migration or genetic variations in drug response, are available to scientists worldwide, with a tube costing about $110. But donors never saw any benefits, Tonko Huenucoy says.

It's a story familiar to others in Chile. "The way research is done nowadays is superconvenient" for scientists, says Constanza Silva Gallardo, a biological anthropologist at Pennsylvania State University, University Park, and a member of the Diaguita Mapochogasta Autonomous Community in Santiago. "There needs to be some sort of pushback to bring effective change."

The proposed constitution could help set the stage, although polls suggest its initial high popularity has recently fallen. But even if it fails, other efforts are ongoing. In March, a mostly Chilean team including Tonko Huenucoy and Silva Gallardo published a paper in *Frontiers in Genetics* urging geneticists to abandon stigmatizing narratives that magnify any genetic differences between Indigenous people and other Chileans. They also called for Chilean universities to develop protocols to incorporate Indigenous voices in designing sampling procedures, drafting informed consent forms, and interpreting results.

In late 2021, this same group launched a program, Ciencia y Comunidades, to improve ethical standards in genomic studies of Indigenous populations in Chile. Last week, they held a workshop at the Pontifical Catholic University of Chile with members of the Aymara, Diaguita, Colla, Chango, Rapa Nui, and Mapuche (including Huilliche and Pehuenche) peoples. Between opening and closing ceremonies involving traditional dances, attendees discussed how research is done, who approves projects, and what genetic data can and cannot say about a person's identity. The effort was modeled after the Summer Internship for Indigenous Peoples in Genomics workshop, an international consortium that explores the ethics of genomics and aims to train Indigenous scientists in the field (*Science*, 28 September 2018, p. 1304).

Cape Town meeting slams 'helicopter research'

By **Cathleen O'Grady**

When researchers from wealthy countries engage in "helicopter research"—field research in poorer countries that extracts data without respectful collaboration—they violate research integrity as well as pose a moral problem, said attendees at last week's World Conference on Research Integrity, held in Cape Town, South Africa.

The conference saw the launch of the Cape Town Statement on equitable research partnerships, which attendees will finalize and submit to an academic journal. Those at the meeting hope their new framing will elevate the issue and help spur systemic solutions, rather than leaving the task of building fair collaborations up to individual researchers.

Researchers in low- and middle-income countries (LMICs) often feel unappreciated when they partner with researchers from wealthier countries, Francis Kombe, co-chair of the African Research Integrity Network and a contributor to the statement, told the conference. Local experts are too often not listed as authors, cannot access data they gathered, and lack the power to steer research to local priorities, studies of the issue have found. All this can affect the quality of research.

Such "scientific colonialism" uses the same tactics as colonialism has historically, Sue Harrison, deputy vice-chancellor for research and internationalization at the University of Cape Town, said at the event. It extracts data instead of raw materials—and undermines and underfunds local infrastructure and skills. This leaves researchers in LMICs without the publications, patents, and skills of their wealthier counterparts.

The Cape Town Statement will offer a guide for how institutions can improve collaborations. For example, funders could set out expectations for equal authorship and data access, says Minal Pathak, a climate researcher at Ahmedabad University in India.

She hopes the statement has impact. "Maybe it's not new. But maybe we need to say it another time."

PHOTO: JAVIER TORRES/AFP/GETTY IMAGES

The goal is to empower communities to demand their rights and to "motivate colleagues to work in a different way," says Constanza de la Fuente, a Chilean ancient DNA researcher at the University of Chicago and a member of Ciencia y Comunidades. "To approach the communities not only saying, 'This is an informed consent form, sign it and give me your sample,' but trying to generate a dialogue with them."

Although they acknowledge that dialogue is needed, some Chilean researchers are wary. In other countries, including Canada, New Zealand, and the United States, Indigenous communities have asked geneticists to delay work, change research questions, keep data private, and not publish results. "One can't take absolutist [attitudes]," such as insisting that scientists must keep all data private, says Lucía Cifuentes, a medical geneticist at the University of Chile (UCh), Santiago. "Science needs creative freedom."

Publishing restrictions would be "censorship," says Ricardo Verdugo, a human population geneticist also at UCh Santiago. But he thinks a new paradigm is needed. Indigenous communities are "the first ones that have the right to have a voice," he says. "What to ask, why ask it, and how I'm going to interpret [and] communicate it, is something that absolutely requires [their] opinion."

For others, now is the moment for drastic measures. "Other scientists might [question] me. But, for me, ethics comes first," says Macarena Fuentes, a human population geneticist at the University of Tarapacá, Arica. "For there to be a transition, extreme changes must occur."

In Puerto Edén, fed up with what they saw as one-sided interactions, the community created a protocol for scientific research within its territory. Scientists must meet with a council to explain their research, what they'll do with the results, and how Puerto Edén will benefit. They must also respect Kawésqar culture, including honoring taboos against visiting sacred places. And they must give something back, whether a simple acknowledgement, a share in any financial rewards, or co-authorship. The plan may be exceptional in Chile at the moment, but many hope it will become the norm in the future.

The protocol isn't a rejection of science, Tonko Huenucoy explains. The community even plans to build a science center and field station to attract research to the community. But they want to make sure that it's done for and with the Kawésqar, she says. So "[our] voices are included from the very beginning." ■

Emiliano Rodríguez Mega is a journalist in Mexico City.

RESEARCH INTEGRITY

Controversial botanist cleared

Report sees "insufficient evidence" of misconduct

By **Charles Piller**

When eight scientists filed a misconduct complaint against prominent botanist Steven Newmaster with the University of Guelph (UG) in June 2021, they thought they had an ironclad case. Their claim that Newmaster, whose work profoundly influenced how dietary supplements are tested and marketed, had made up or plagiarized data in three papers was "an entirely credible and well-founded allegation," says cosignatory Kenneth Thompson, a postdoc at Stanford University.

But an investigative committee at UG disagrees. Newmaster "displayed a pattern of poor judgement," his conduct was "suspicious," and there were "many shortcomings" in his work, panel chairman John Walsh, a business professor at UG, wrote in a 1 June letter to the complainants. But there was "insufficient evidence" to find Newmaster guilty of misconduct.

"Given the evidence of data falsification that was assembled, I was very surprised at the conclusion," says evolutionary biologist Paul Hebert, a co-signatory to the complaint. Hebert directs UG's Centre for Biodiversity Genomics and pioneered DNA barcoding, a technique for identifying organisms from short snippets of DNA that is central to Newmaster's work.

> **"I was very surprised at the conclusion."**
> **Paul Hebert**, University of Guelph

But Thomas Braukmann, a former postdoc at the UG center who is now at Public Health Ontario, says the findings are "disappointing but not surprising." "I don't think [UG] took the allegations seriously or put the committee together in good faith," says Braukmann, who was not involved in the case but studied Newmaster's papers at *Science*'s request. "We need a better system in Canada to handle misconduct concerns." Walsh and UG did not respond to requests for comment.

Newmaster made headlines with a 2013 *BMC Medicine* study reporting that many herbal supplements didn't contain the labeled ingredients and some had toxic contaminants. The paper propelled him to global fame as a testing expert. His own companies and a nonprofit group at UG raised millions of dollars by certifying supplements, cannabis, and other comestibles.

Thompson was the first to level accusations against Newmaster, in 2020. He claimed Newmaster made up and plagiarized the data for a study testing the ability of DNA barcoding to identify species in a Canadian forest, which they published together in 2014, when Thompson was a UG undergraduate. After the university dismissed his complaint, the group of eight sent a more elaborate complaint letter. It fingered not only Thompson's paper—which the journal *Biodiversity and Conservation* retracted in October 2021—but also the supplements study and a 2013 paper using DNA barcoding to study woodland caribou diets. Newmaster's co-authors have requested retractions of those articles as well.

For now, *BMC Medicine* is investigating the allegations about the supplements article and the Canadian *Journal of Forest Research* has added an Expression of Concern to the caribou paper. Newmaster has denied the charges. "I have never engaged in any unethical activity or academic misconduct," he wrote in an official reply obtained by *Science*.

An investigation by *Science* (3 February, p. 484) revealed many other cases in which Newmaster appeared to manipulate or fabricate data, plagiarize, and invent elements of his academic record. He did not respond to requests for comment for the *Science* story, and the panel did not address the issues it raised. UG could take up to several months to issue a final decision.

Newmaster's accusers had been worried the investigation might not be rigorous, given the panel's lack of expertise in genomics. In addition to Walsh, it included Jeff Wichtel, dean of UG's veterinary college, and Cynthia Fekken, a psychologist from Queen's University. Walsh's letter says they relied on an independent expert witness, whom he did not name. The letter notes that a "key factor" in the panel's finding that it was impossible to "definitely establish" misconduct was the "absence of records, including raw data."

That's a particularly frustrating argument, Thompson says. "That was the essence of our complaint. We knew they wouldn't be able to find records. Our complaint alleged that Prof. Newmaster falsified his work and never had the data to back it up." ■

This story was supported by the *Science* Fund for Investigative Reporting.

U.S. RESEARCH FUNDING

At NSF, what price geographic diversity?

U.S. innovation bills present Congress with dueling visions for funding have-not states

By **Jeffrey Mervis**

Top research universities in a handful of U.S. states conduct the majority of research funded by the National Science Foundation (NSF), whereas institutions in half the country receive only crumbs.

The Senate and House of Representatives both want to redress the imbalance, but lawmakers in the two chambers have proposed different solutions. Those dueling visions must be reconciled in negotiations now underway to finalize a massive bill aimed at bolstering U.S. competitiveness with China in research and high-tech manufacturing that is a top priority for President Joe Biden.

The U.S. Senate thinks the solution is to mandate an eightfold increase in an existing initiative—the Established Program to Stimulate Competitive Research (EPSCoR)—that serves only the have-not jurisdictions (25 states, Puerto Rico, Guam, and the Virgin Islands). To finance that huge growth in EPSCoR, NSF would need to shrink core research programs in which the money is awarded competitively.

More than 200 universities and more than 100 members of Congress have launched a last-minute fight to remove the Senate language from the final bill. "Arbitrarily walling off a sizable percentage of a science agency's budget from a sizable majority of the country's research institutions would fundamentally reduce the entire nation's scientific capacity," warned 18 senators and 78 House members in a 24 May letter that echoes a 2 April plea to lawmakers from a coalition of institutions in non-ESPCoR states. Curtailing existing NSF programs, they say, would harm many less research-intensive institutions located outside of have-not states.

The critics prefer the House approach. It has proposed new programs to help poorly funded institutions in every state. But the Senate version may have the upper hand in negotiations. It is tucked into the U.S. Innovation and Competition Act (USICA), which passed in June 2021 with strong bipartisan support. The House version, the America COMPETES Act, was approved by a narrow margin in February with only Democratic members in favor.

NSF says roughly 13% of its research funds now go to institutions in EPSCoR states. But only about one-fifth of that comes directly from EPSCoR. The bulk is awarded competitively through NSF's regular research and training programs. The Senate bill would allocate 20% of NSF's overall budget, currently $8.8 billion, to EPSCoR. If the rule were in place now, it would hike EPSCoR's budget from $215 million to roughly $1.75 billion.

Last month, NSF Director Sethuraman Panchanathan told a Senate spending panel he has "an aspirational goal" of sending 20% of NSF's research budget to have-not states. But that approach to increasing geographic diversity didn't satisfy the panel's chair, Senator Jean Shaheen (D–NH), who instructed her Senate colleagues negotiating the bill "to hold tight to the 20% requirement" for EPSCoR itself.



Senator Maria Cantwell (D–WA, right) is trying to steer a massive bill through Congress that contains a controversial provision for geographic diversity in research funding crafted by Senator Roger Wicker (R–MS).

PHOTO: SUSAN WALSH/AP PHOTO/BLOOMBERG/GETTY IMAGES

Shaheen was one of 33 Senators and 26 House members, all from EPSCoR states, who last fall signed a letter to negotiators arguing for preserving the Senate language. "If the United States is going to stay a step ahead of China, we need to promote the scientific talent, expertise, and capabilities found throughout America, not just in a handful of states," asserts Senator Roger Wicker (R–MS), who spearheaded the letter and crafted the EPSCoR provision in the Senate bill.

Senator Maria Cantwell (D–WA), co-chair of the 107-member negotiating committee and chair of the Senate commerce committee that oversees NSF, has not taken a position on the set-aside but says the geographic imbalance "is something that we have to address." And last month, at the conference committee's first meeting, she praised Wicker's "passion for spending federal research dollars in areas defined by EPSCoR."

In contrast, the House's COMPETES Act applauds EPSCoR for "improving research capacity and competitiveness" but gives the program no specific set-aside. Instead, it would create several NSF programs aimed at fostering greater geographic diversity in research.

One would let institutions not in the top 100 recipients of federal research dollars—in 2020 that meant receiving $255 million or more—compete for money to conduct research, recruit faculty, offer student stipends, and carry out "other activities necessary to build research capacity." Another would fund joint projects between research powerhouses and "emerging research institutions"—defined as those receiving less than $35 million annually in federal research funding.

Despite a budget that has tripled since 2001, EPSCoR has not reshaped the geography of NSF's spending. In 2020, just five states—California, Massachusetts, New York, Texas, and Maryland—received nearly 40% of NSF's research dollars, whereas the bottom five—Vermont, West Virginia, Wyoming, and North and South Dakota—together got less than 1%.

The Senate's views on EPSCoR carry weight because the final bill will need the backing of most of the 19 Senate Republicans who voted for USICA. Biden has even renamed the proposed legislation, calling it the Bipartisan Innovation Act in hopes of retaining Republican support.

Democratic leaders have asked conferees to reach agreement this month. With so much to like in the bills, science advocates also want to see the legislation enacted—but not at the cost of disrupting how NSF funds research. ■



Volcanic glass, like that found near Iceland's Blue Lagoon, can help knit RNA letters into long strands.

PHOTO: SURANGA WEERATUNA/ALAMY STOCK PHOTO

BIOCHEMISTRY

Did black volcanic rock help spark early life?

Quenched lava may have helped form long RNA strands vital to primordial organisms

By **Robert F. Service**

When life emerged, it did so quickly. Fossils suggest microbes were present 3.7 billion years ago, just a few hundred million years after the 4.5-billion-year-old planet had cooled enough to support biochemistry. Many researchers think the hereditary material for these first organisms was RNA. Although not as complex as DNA, RNA would still be difficult to forge into the long strands needed to convey genetic information, raising the question of how it could have spontaneously formed.

Now, researchers may have an answer. In lab experiments, they show how rocks called basaltic glasses help individual RNA letters, known as nucleoside triphosphates, link into strands up to 200 letters long. The glasses would have been abundant in the fire and brimstone of early Earth; they are created when lava is quenched in air or water or when the melted rock created in asteroid strikes cools off rapidly.

The result has divided top origin-of-life researchers. "This seems to be a wonderful story that finally explains how the nucleoside triphosphates react with each other to give RNA strands," says Thomas Carell, a chemist at the Ludwig Maximilian University of Munich. But Jack Szostak, an RNA expert at Harvard University, says he won't believe the result until the research team better characterizes the RNA strands.

Origin-of-life researchers are fond of a primordial "RNA world" because the molecule can carry out two distinct processes vital for life. Like DNA, it's made up of four chemical letters that can carry genetic information. And like proteins, RNA can catalyze chemical reactions needed for life.

But RNA also brings headaches. No one has found a set of plausible prebiotic conditions that would cause hundreds of RNA letters—each of them complex molecules—to link into strands long enough to support the complex chemistry needed to ignite evolution.

Stephen Mojzsis, a geologist now at the Research Centre for Astronomy and Earth Sciences of the Hungarian Academy of Sciences, wondered whether basaltic glasses played a role. They are rich in metals such as magnesium and iron that promote many chemical reactions. And, he says, "Basaltic glass was everywhere on Earth at the time."

He sent samples of five different basaltic glasses to the Foundation for Applied Molecular Evolution. There, Elisa Biondi, a molecular biologist, and her colleagues ground each sample into a fine powder, sterilized it, and mixed it with a solution of nucleoside triphosphates. Without a glass powder present, the RNA letters failed to link up. But when mixed with the glass powders, the molecules joined into long strands, some hundreds of letters long, the researchers report last week in *Astrobiology*. No heat or light was needed. "All we had to do was wait," Biondi says. Small RNA strands formed after just a day, but strands kept growing for months. "The beauty of this model is its simplicity," says Jan Špaček, a molecular biologist at Firebird Biomolecular Sciences. "Mix the ingredients, wait for a few days, and detect the RNA."

Still, the results leave questions unanswered. One is how the nucleoside triphosphates could have arisen in the first place. Biondi's colleague Steven Benner says recent research shows how the same basaltic glasses could have promoted the formation and stabilization of the individual RNA letters.

A bigger issue, Szostak says, is the shape of the RNA strands. In modern cells, enzymes ensure most RNAs grow into linear chains. But RNA can also bind in complex branching patterns. Szostak wants the researchers to report which type of RNA the basaltic glasses created. "I find it very frustrating that the authors have made an interesting initial finding but then decided to go with the hype rather than the science," he says.

Biondi admits her team's experiment almost certainly produces a small amount of RNA branching. However, she notes that some branched RNAs exist in organisms today, and related structures may have been present at life's dawn. She also says other tests the group performed confirm the presence of long strands with connections that most likely mean they are linear. "It's a healthy debate," says Dieter Braun, an origin-of-life chemist at Ludwig Maximilian. "It will trigger the next round of experiments." ■

FEATURES



Poached to extinction for their mystique and beauty, Spix's macaws (opposite page) will reenter the wild from this aviary in Brazil.

A WILD HOPE

Two decades after it disappeared in nature, the stunning blue Spix's macaw will be reintroduced to its forest home

By **Kai Kupferschmidt**, *in Curaçá, Brazil*

In 1995, conservationists and scientists embarked on a desperate attempt to save the world's rarest bird, a blue-gray parrot called the Spix's macaw. The bird had scarcely been spotted since scientists first described it in the early 19th century, and it had taken on an aura of mystery, making it irresistible to parrot lovers—and to poachers. "For well over a century we just had this very, very weak information that there was this kind of mythical, rather beautiful blue bird," says Nigel Collar, a conservationist at BirdLife International. By the mid-1990s only a single individual remained alive in the wild, close to this dusty, small town in northeastern Brazil.

From DNA in molted feathers, researchers in the United Kingdom confirmed that the last wild bird was a male. At the time, fewer than three dozen birds were known to be held in collections and zoos around the world, and a decision was made to release a single female in hopes the birds would pair and produce offspring. The female was released close to where the male lived and seemed to quickly adapt to her new life, eating wild food and avoiding an attack by a falcon. She grew stronger by the day, flying farther and farther, and after little more than 2 months had paired with the male. Two weeks later, she mysteriously disappeared.

Years later, a local man said he had found the bird dead below a power line. "If that's really true, then that is just incredibly bad luck," Collar says. It is almost unheard of for parrots to hit power cables, he says, and in reality she might have been taken by poachers. "The world of Spix's macaw is full of very, very great uncertainties and a lot of people who say a lot of things that they don't necessarily really mean." The wild male vanished

PHOTOS: MARTIN GUTH; (OPPOSITE PAGE) PATRICK PLEUL/PIC



a few years later, and the Spix's fate seemed sealed—another species lost.

Now, conservationists are attempting to undo that fate. On 11 June, more than a quarter-century after the female flew into oblivion, they plan to release eight Spix's macaws from captivity into the wild. Twelve more are supposed to follow at the end of the year and still more in the years to come. If everything goes according to plan, these birds will be the vanguard of a new population of Spix's macaws in their natural habitat. The project, long hampered by infighting and overshadowed by controversy, had to overcome significant scientific hurdles to even come this far. But the biggest challenge still lies ahead.

"The Spix's project is unique in that they are reintroducing a species back into the wild that is currently extinct, has been extinct in the wild for over 2 decades," says Thomas White, a wildlife biologist at the U.S. Fish and Wildlife Service and a technical adviser to the project. "There's very few reintroduction programs around the world that have done something like that, none with parrots or macaws."

Few reintroductions of birds have been successful, and none was as ambitious as this one, says George Amato, a conservation biologist at the American Museum of Natural History. Yet for the Spix's it has to be tried, he says. "I hope it works, because we really have no other alternatives."



In the wild, Spix's macaws nest in the hollows of large caraibeira trees, which grow along streams in the dry forest of northeastern Brazil.

PHOTO: LUIZ CLAUDIO MARIGO/MINDEN PICTURES

THE NATURAL HOME of the Spix's macaw is in the caatinga, a tropical dry forest in northeastern Brazil that covers 10% of the country. In the rainy season, which lasts for about 2 months, everything appears lush and green. But the rest of the year plants here compete in shades of gray and white—caatinga means "white forest" in the Indigenous Tupi language. It is here that the Spix's macaws once nested in the hollows of old caraibeira trees growing along the creeks that cut through the caatinga, feeding on seeds and nuts.

It is impossible to know how many Spix's macaws existed in the past. By the time Western science discovered the bird, humans had already started to parcel large parts of the caatinga into ranches. In 1819, German naturalist Johann Baptist von Spix spotted the parrot on an expedition to the interior of Brazil. Spix noted that the bird appeared to be "very rare"—then shot it and brought it home to Munich, setting the tone for humanity's relationship with this striking bird going forward.

As the human footprint increased in the caatinga, the bird became even rarer. Tragically, this only made it more coveted by parrot collectors, who were willing to pay tens of thousands of dollars for a single bird. "The rarer it was, the more it became a kind of status symbol," Collar says. The bird became something akin to the exceedingly rare blue Mauritius stamp coveted by philatelists, says Roland Wirth, a conservationist at the Zoological Society for the Conservation of Species and Populations. "The very wealthy, very passionate collectors really wanted to have one, and they would do almost anything to do so."

By the beginning of 1987, only three Spix's macaws were known to survive in the wild, and by the end of that year, poachers had taken two of them. After the plan to pair the last male with a captive bird failed in 1995, the male remained with a female of a different species, an Illiger's macaw, until he, too, disappeared in October 2000. The International Union for Conservation of Nature officially declared the Spix's macaw extinct in the wild in 2019, exactly 200 years after Spix had described it.

Even then, the bird retained its hold on the popular imagination. The story of the last lone male inspired songs—including one written from the perspective of the Illiger's female waiting in vain for his return—and two animated movies that together earned $1 billion.

ON A HOT MORNING in February, Martin Guth, a bald and burly German businessman and parrot collector, stood in the spot where the Spix's will begin its new life in the wild. The nongovernmental organization (NGO) he founded, the Association for the Conservation of Threatened Parrots (ACTP), has taken on the challenge of bringing the bird back to the caatinga. ACTP, which houses more than 170 Spix's macaws in Tasdorf, near Berlin, built a facility just a few hundred meters from where Guth is standing and, in March 2020, flew 52 macaws to Brazil by private jet to take up residence there. In 2021, three chicks hatched at the facility, the first Spix's born in the bird's original home in more than 30 years.

But that morning, Guth was angry. Nearby, workers were busy constructing a huge U-shaped aviary where the birds will be able to fly longer distances than they can in their small cages inside the main facility. It was running behind schedule. "Even on the way here, the guy still said everything was finished," Guth grumbled. He was convinced that a rival who was previously involved in the Spix's project had something to do with the delay. The Spix's project may have high-minded goals, but its history is replete with jealousies and backbiting.

The idea of breeding Spix's macaws in captivity and reintroducing them to the wild began long before Guth's involvement, and even before the lone wild male had disappeared. In 1990, conservationists formed a committee to oversee a reintroduction

program. That meant building up an adequate captive population, which proved to be a complicated and controversial process.

At first conservationists only knew of a few captive birds—and owners were reluctant to come forward, because the export of wildlife had been illegal in Brazil since 1967. But the Brazilian government agreed to grant amnesty to owners if their birds joined the breeding program, and "one by one, people came out and admitted they had Spix's macaws," says Wolfgang Kiessling, a businessman who founded and runs Loro Parque, a private zoo on the island of Tenerife that held some Spix's macaws for many years.

Still, by 1996 only 39 captive birds were known around the world. Making matters worse, most of them were closely related. Only nine of the birds had come from the wild, and 21 of the remaining 30 were offspring from a single pair in the Philippines, raising concerns about inbreeding. For the Spix's to have any future, birds from different collectors needed to be brought together to breed, but arguments over who would send a bird to whom under what conditions kept derailing the plans. "The rarer the animal, the more politics is involved," says Cristina Miyaki, a bird geneticist and a member of the advisory committee of the Spix's project. In contrast to the spirit of cooperation required for a successful recovery effort, Collar wrote in 1992, "ownership is a matter of jealousy, prestige and possessiveness that is fundamentally different in psychological origin."

Meanwhile, the constellation of owners kept changing. Starting in 2000, Sheikh Saoud Bin Mohammed Bin Ali Al-Thani of Qatar bought dozens of Spix's to keep at Al Wabra, his private wildlife preserve. In time, he came to own the vast majority of all known Spix's macaws in the world.

Guth entered the scene in 2005, beating out the sheikh to buy from a private Swiss owner three Spix's macaws that had not previously been part of the breeding program. "The three birds he had were the most important ones, because they could improve the genetics of the population," says Camile Lugarini, a veterinarian at the Chico Mendes Institute for Biodiversity Conservation (ICMBio), who leads the Spix's macaw project for the Brazilian Ministry of the Environment.

In May 2012, a meeting in Brazil's capital, Brasília, brought together representatives of all the important stakeholders. It was testy. One participant argued that Guth should have no part in the project because he had served a prison sentence and, this person claimed, had sold endangered birds illegally, in violation of the Convention on International Trade in Endangered Species of Wild Fauna and Flora. (Guth says that like other breeders and NGOs, ACTP sells some birds legally, but has never sold Spix's or other highly endangered birds, and that his offenses were committed decades ago and have nothing to do with the current project.) Tim Bouts, a veterinarian who was then the curator at Al Wabra and attended the meeting, says he spoke in defense of Guth, who was not present: "Let's be honest, this table here is full of criminals. Every single Spix's that came into captivity was illegal."

The meeting ended with no agreement.

Guth has pressed ahead, even as some have questioned his motives and methods, pointing to the lack of transparency around ACTP and its sources of funding. Guth says some donors prefer to remain anonymous and that he is trying to avoid the disputes over funding and credit that doomed the project in the past. "Yes, we are doing things differently," he says. "It certainly didn't work the way they tried it before."

Even some people who say they are intimidated by Guth acknowledge the effectiveness of his pushing, bullying, and cajoling. "He is a bit of a bulldozer," Wirth says. "But he gets things done." When the sheikh died suddenly in 2014 and the future of his Spix's macaws was in doubt, Guth stepped in through his NGO to bring the birds from Qatar to Tasdorf. In June 2018, Guth and Brazil's environment minister signed a memorandum of understanding in Berlin to build the facility in Brazil, transfer birds, and reintroduce them. (ICMBio and the Pairi Daiza Foundation were also signatories.)

"I wasn't born as a conservationist," Guth says. But as he became involved in the reintroduction effort, he grew determined to prove his critics wrong. "They said, 'He won't be able to breed the birds.' I did. They said, 'He won't send any birds to Brazil.' I did. They said, 'He won't reintroduce the birds.' We are doing that."

He has put himself in an interesting position, Collar says. "He is the one now who can go down in history as the person who saved the Spix's macaw. Or if he really messes up, then he goes down in history as the person who made it go extinct."

WHILE OWNERS WERE FIGHTING over control and credit, conservationists and researchers were fighting to save the species. Whe ornithologist Cromwell Purchase went t Al Wabra in 2010 to head its Spix's maca program, he was told the species was "o the fence." At the time, 54 of 71 bird known worldwide were in Qatar, an the captive population faced twi threats: disease and a low birth rate

The major disease threatening cap tive Spix's was proventricular dilata tion disease, which affects the nerve in parrots' gastrointestinal tract an causes them to slowly waste away. common scourge of pet birds, it ha been known since the 1970s, but it cause was completely unclear. Then in 2008, researchers identified novel virus in the brains of affecte birds: a type of bornavirus, a grou known to cause brain disease i horses and sheep.

"We tested all known Spix's in th world for this virus," says Michae Lierz, a veterinarian at the Justu Liebig University Giessen. In Qata a traffic light system was implemented with infected birds deemed "red" an separated from the others. This eventuall eliminated the threat of avian bornaviru to the Spix's population.

The other problem was reproduction Only a few pairs were producing chicks At first a decision was made to keep the reproducing. "The goal was to produce a many animals as possible to keep the specie from going completely extinct," Lierz says. Over time the focus shifted to making better matches, in order to preserve the Spix's genetic diversity and, therefore, its chances of survival. But birds with diverse genetics wouldn't necessarily form a pair. "Parrots are monogamous and very choosy," Lierz says.

So, veterinarians at Al Wabra considered artificial insemination. For many birds, including chickens, pigeons, and birds of prey, this is fairly straightforward, Lierz says. The technique involves massaging a male's cloaca from the outside with the thumb. ("A short and smooth thumbnail is

Lost ground

The Spix's macaw's native home is the caatinga, a dry tropical forest that is leafless most of the year. The forest has dwindled because of ranching, another obstacle facing the effort to reintroduce the macaw.



CREDITS: (MAP) N. DESAI/*SCIENCE*; (DATA) TERRABRASILIS/NATIONAL INSTITUTE FOR SPACE RESEARCH

advantageous for performing cloacal massage and protects the bird from accidental injury," one paper notes.) But this technique does not work on most large parrots. Around 2010, Lierz and his colleague Daniel Neumann developed a new method: inserting a small probe into the cloaca to deliver a weak electric current that stimulates a male bird to release sperm. "As kids we used to hold these 9-volt batteries to our tongues and it tingled, that's roughly how you have to imagine this," Lierz says.

With artificial insemination, the researchers could finally pair birds according to their genetics. But the timing was crucial: Females usually lay two or three eggs and the moment one egg is laid is the right time to inseminate the next one. Purchase says he and Neumann spent hours watching female Spix's macaws on video monitors. "As soon as we see the egg, we're out and we go from aviary to aviary and we catch the male that we want, male No. 1 on the list. We try and collect semen from him, and if we don't get enough … then we go to male No. 2," Purchase says. In May 2013, the first artificially inseminated Spix's macaw chicks hatched. More followed. "That's what got us out of the genetic bottleneck," Bouts says.

THE MORNING AFTER GUTH was fuming about the aviary delay, Purchase walked into a large room at the facility carrying a gray plastic cage in each hand. He set them down on the tiled floor, opened the door of one, reached inside with a dark towel, and enfolded what was inside. Kneeling on the floor, he delicately unwrapped the towel. A gray head emerged first, then turquoise feathers covering the parrot's belly, and finally the rich blue of its back and tail.

Purchase carried the bird over to Francois Le Grange, a veterinarian, who began to examine it—a final check before it would join the other candidates for release in the not-quite-finished outdoor aviary. The bird's outraged "ca-á ca-á" echoed off the walls as Le Grange plucked a feather from beneath its wing. Then he listened to its heartbeat with a children's stethoscope. He swabbed the mouth and the cloaca and finally drew some blood from a vein in the neck.

The swabs would be tested for pathogens that might pose a risk to other animals after the birds are released. But the team is much more worried about the dangers these parrots themselves will face in the wild. After generations in captivity, their instincts for navigating and finding food have weakened, White says. There are predators, too, including opossums, snakes, and birds of prey. And, of course, humans—the species that drove the bird to extinction in the first place.

Together these challenges doomed some earlier reintroduction programs. One of the highest profile examples was an attempt to bring the thick-billed parrot back to Arizona, Amato says. This brightly colored bird still lives in Mexico but has been hunted to extinction in the United States. Between 1986 and 1993, 88 of them (mostly confiscated birds originally trapped illegally in Mexico, but also some captive-bred birds) were released in the Chiricahua Mountains in Ari-



Johann Baptist von Spix first described and painted the macaw in an 1824 publication.

zona. Many were killed by hawks or cats or starved to death. After 2 months, only about two-thirds of the wild-caught birds survived. But the captive-bred birds did much worse, as a paper noted in 1994: "Almost all individuals have been lost within a few days of release as a result of substantial deficiencies in basic survival skills." The program was abandoned in 1993, and the last time a thick-billed parrot was spotted in Arizona was in 1995. "The release program was a failure, even though a lot of money and effort was spent on it," Amato says. "After that, many biologists felt that release programs for parrots generally were unlikely to be successful."

Yet Amato notes some hopeful counterexamples: the feral populations of escaped parrots that thrive in many parts of the world, including London and New York City. "These are like accidental reintroductions that worked," he says. Some recent planned reintroductions have also had positive results, White says, including one he was involved in: the reintroduction of Puerto Rican parrots to El Yunque National Forest after they were wiped out by Hurricane Maria in 2017. Since 2020, 75 captive-reared animals have been released in the forest, which now hosts 34 birds. Four new nests were spotted this year, White says. "This was a true reintroduction and it has been very successful."

THAT NIGHT, as darkness descended over the caatinga, Lugarini headed out with a colleague to a creek near the facility. Wearing leather gaiters to protect against snakes, she followed the mostly dry creek bed, moving as quietly as possible. She stopped in front of a caraibeira tree, where a pair o Illiger's macaws had made their nest.

Illiger's macaws, also known as blue winged macaws, play an important role i the plan to bring back the Spix's. Illiger' are more common and inhabit a larger are than the Spix's macaws, but in the caating the two birds' lifestyles overlap. Both nest i hollows in caraibeira trees and feed on th same fruits and nuts. When the eight Spix' are released, eight Illiger's macaws take from the wild will be released with them The team hopes this mixed flock will joi up with wild Illiger's in the caatinga, allow ing the Spix's macaws to benefit from thei knowledge of how to avoid predators, fin food, and navigate.

The team had already collected seve Illiger's, and Lugarini had come for th eighth. Her headlamp casting a red glow she grasped a cord looped around a branc high in the caraibeira tree, fixed a rope t it and then used the cord to pull the rop over the branch and back down. Lookin up, she sighed with apprehension at th sight of bats circling the tree. "That's wors than the snakes," she said. Yet she slowl ascended the rope, the red light markin her progress. Ten meters up she reached th nesting hollow and looked inside. No birds The Illiger's macaws that had been nestin here were gone.

One clear lesson from previous reintro ductions is that releasing more animals i better. That's because a bigger group ca work together to spot dangers and find food. Finding a suitable mate is easier, too. For highly social species like macaws, numbers are especially important. "Let's say you release 20 individuals and they all go 20 different directions, well then you haven't reestablished a population," White says. "They need to live in a group." Combining captive Spix's and wild Illiger's thus solves two problems, White says. "We can actually increase the flock size without extra Spix's … while using a native species which knows the habitat, knows the area, that can function as mentors."

ILLUSTRATION: DAVID TIPLING PHOTO LIBRARY/ALAMY STOCK PHOTO





Veterinarian Francois Le Grange (top right) and animal keeper Sebastian Laurisch examine a Spix's macaw at the breeding station in Germany where these three chicks (bottom) were born.

PHOTOS: PATRICK PLEUL/PICTURE ALLIANCE VIA GETTY IMAGES (2)

Releasing birds of the right age can help keep them from scattering. Spix's macaws start to reproduce around age 4 and then tend to return to the same nesting site year after year. "The sooner that those released macaws start reproducing, the sooner they become anchored to that site," White says. "So you want to have birds that are entering or at reproductive age." Providing supplementary food and nest boxes may also encourage the birds to remain close to the release site.

When White and other researchers reviewed 47 releases of captive parrots into the wild, they found that the single biggest threat to success was predation. To reduce this risk, Purchase put metal bands around trees with nest hollows or nest boxes to keep predators like opossums from climbing the trees. To avoid tipping off would-be poachers, he put decoy bands around trees without nests, as well. The birds will also wear tracking collars.

After a long day, Lugarini headed back to her hotel in Curaçá. As the four-wheel drive vehicle bounced over the dusty road, goats scattered and closed wooden gates slowed her progress. It was a reminder that the Spix's natural habitat barely exists anymore. A restoration project is ongoing but has been hampered by the lack of knowledge of this little-studied biome and by its location in one of the poorest regions of Brazil, where the goats that provide a lifeline for the local population have devoured much of the natural vegetation.

"In the beginning I did sometimes think, 'Why are we putting all this effort into bringing back one species that is extinct when there are so many other species that we could still save from extinction?'" Lugarini says. "But you have to remember that this flagship species helps us preserve and restore the caatinga, and that helps many other species, too."

Curaçá is home to about 30,00 inhabitants—and many homages to th Spix's. Next to the gas station is the Spi hotel. The theater, restored with mone from the Spix project, is bright blue. Th city's flag in front of the town hall include a Spix's macaw, though Lugarini notes "the got it wrong": The bird has the yellow mark ings around the eyes and next to the bea that are typical of the Lear's macaw, anothe threatened macaw that lives not far away.

One resident, Fernando Ferreira, wrot the song about the lovesick Illiger's macaw Wearing shorts and a T-shirt, his gray hai swept back in a ponytail, Ferreira sat dow with a guitar and sang another song h wrote about the Spix's macaw, known her as *ararinha azul*, or little blue macaw: "M wish is to see you fly, my wish is to see yo come back," he sang. On the afternoon o 11 June, Ferreira will perform this song a a ceremony at the theater. There will be video, speeches, and a press conference Earlier that day, in front of a small group o people, Purchase will open the door of th aviary to release the birds.

For those who have worked toward thi for years, it will be a moment of joy an apprehension. "It will feel like a weigh off my shoulders, probably," Purchase says But then comes the next weight—worryin about their survival. There is an element o guilt, Miyaki says, because humans drov the Spix's to extinction. "We owe it to th species, for it to go back to the wild." But th experience of 1995 still casts a shadow, she says. "The frustration after the first release of that female was so big," she says. "I try to be optimistic, but I'm very anxious."

The project estimates that between one-third and two-thirds of the birds will be lost in the first year. If the losses are higher, the birds may be taken back in. "You try to make sure that you have covered all of the bases and thought about as many possible options and outcomes as possible," White says. "But the day you release those birds, the day they leave that cage, a lot of things are no longer within your control." ■

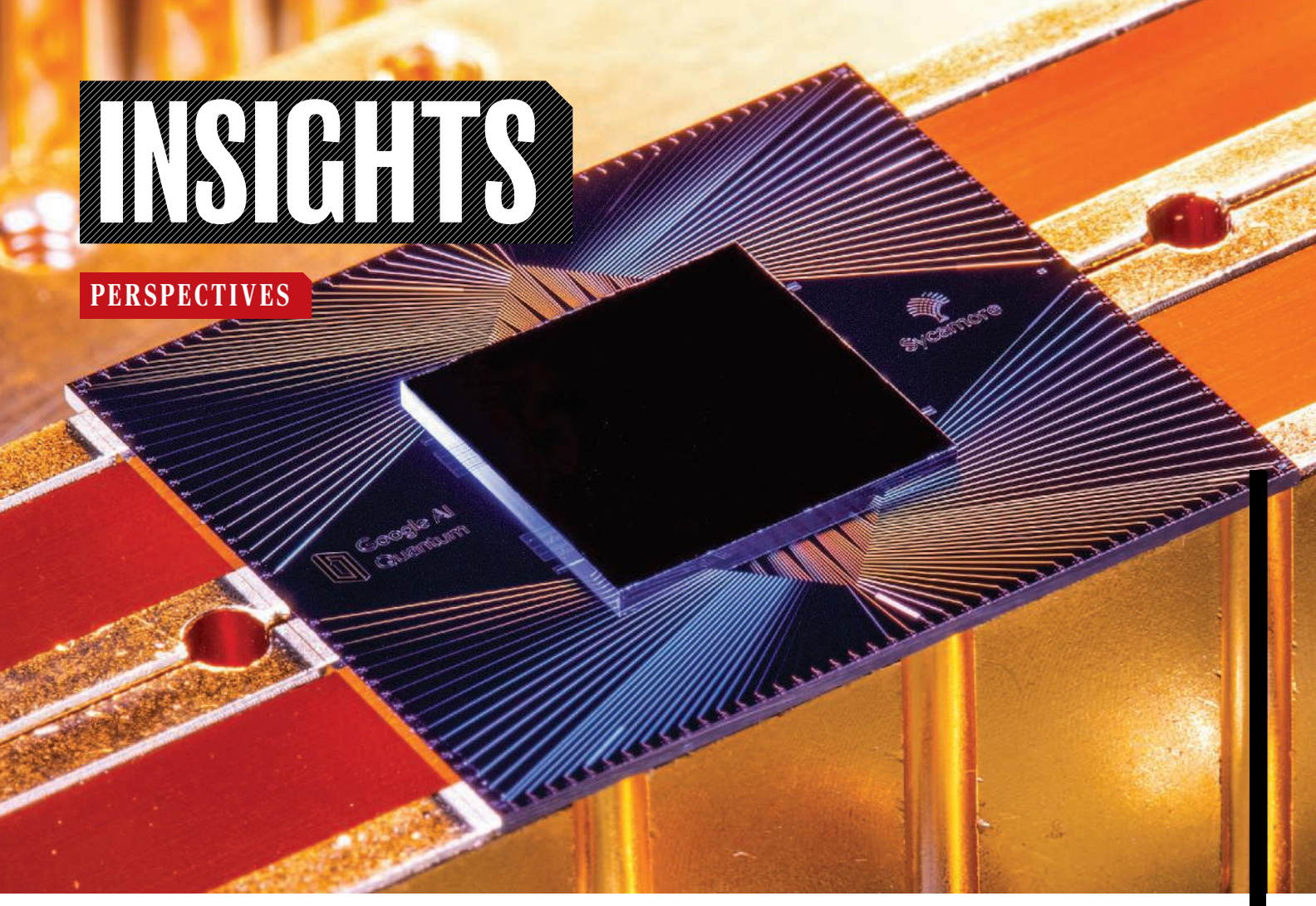


PHOTO: ERIK LUCERO/GOOGLE

QUANTUM COMPUTERS

Quantum learning unravels quantum system

A quantum computer has a decisive advantage in analyzing quantum experiment results

By **Vedran Dunjko**

In the early 1980s, American physicist Richard Feynman proposed that machines can exploit quantum phenomena to perform otherwise intractable computations. The kind of computation he envisioned was broadly about simulating the properties of a quantum system given its classical description. In the decades that followed, researchers have identified numerous other problems that in theory can only be solved within a reasonable time frame by using such a quantum computer. However, a quantum computer that can exercise this advantage over a classical computer does not exist yet. A recent hope is that near-term quantum computers may be used as a new type of machine learning device that offers an edge in analyzing data from quantum experiments. On page 1182 of this issue, Huang *et al.* (*1*) present an experimental realization of a quantum learning algorithm that has a provable advantage over its conventional counterpart while being within the reach of today's quantum computers.

The signature capacity of a quantum computer is the ability to predict the behavior of a many-particle quantum system when given its initial condition. With the development of machine learning methods, even more complex questions can be asked. Machine learning can enable such prediction even without full knowledge of the system, but with merely having access to previous experimental data. The task can be thought of as a two-stage process. The computer is fed a dataset that stems from some previous experiments over a quantum system. Then, the classical or quantum computer will have to predict the future of the system under slightly different settings. Intuitively, such data analysis may inherit the so-called quantum-classical performance gap, as described by Feynman: that the computation of the future state of the system, given its full description, is intractable for classical but feasible for quantum computers.

However, unexpectedly, the inclusion of training data could close this gap. Classical machine learning can sometimes predict properties of complex quantum systems (*2*), so it is unclear whether quantum computers hold an edge in this setup. Huang *et al.* propose an approach that will give quantum computers a decisive edge: by leveraging the quantum computer's ability to process "quantum data," raw quantum states that result from a quantum experiment and not

Applied Quantum Algorithms group, Leiden Institute of Advanced Computer Science; Lorentz Institute for Theoretical Physics, Leiden University City, Netherlands. Email:v.dunjko@liacs.leidenuniv.nl

By harnessing the power of the Google Sycamore processor (pictured here), Huang *et al*. showcase the exponential advantages offered by quantum computers for analyzing data from quantum experiments.

mere classical information. The quantum data can be used by the quantum computer to predict future results while requiring far fewer experiments.

One cannot input quantum data into a classical computer. Intuitively, one may think that this would give the quantum device a straightforward advantage. But the actual scenario is more nuanced. For a classical setup, the quantum state of an experiment can be measured and used as input, with each measurement freely chosen by the classical learning algorithm. Because the classical computer can arbitrarily choose when to measure each of the experiments, then at least in principle, all the information encoded in quantum states can be accessible. For a quantum setup, the quantum computer provides a minimal but key additional capacity: a small quantum memory that enables joint measurements on two copies of quantum data.

So in both cases, all the quantum data are converted to classical information before calculation, but in slightly different manners. Joint measurements—used in the quantum-enhanced scenario—unravel correlated properties of two separate quantum systems. This fundamentally exploits quantum entanglement (when the quantum states of two or more objects are intertwined with each other) and cannot be substituted by pairs of individual measurements. Since the early days of quantum information theory, it has been known that joint measurements can help distinguish quantum states, even when the states are uncorrelated (*3*). But until recently, it was not clear just how large an advantage this exploit can give quantum computers over their classical counterparts.

Building from the research line on so-called shadow tomography (*4*–*6*), Huang *et al*. argued that joint measurements lead to substantial advantages for learning about quantum systems. Namely, the quantum-enhanced strategy is exponentially more economical in terms of the number of quantum experiments needed for predicting the outcomes of just two measurements (*6*). The authors demonstrated the advantages of a quantum learning experiment using the Google Sycamore processor. The natural scenario of quantum data learning involves a "transducer" that transports the quantum state of results from an experiment into the quantum computer. Their experiment was simulated in the same quantum processor that analyzes the data, in a lab-on-a-chip setting. Once the quantum state is prepared, it is analyzed with classical and quantum-enhanced methods.

For the optimal predictions, the exact joint measurements may be known, at least in idealized settings. However, in the real experiment, the state preparation is imperfect, as is the measurement performed. To counteract this, the quantum processing is supplemented with classical machine learning to extract the strongest signals in the presence of experimental errors. This classical-quantum hybrid approach demonstrates advantages in our capacity to learn various fundamental properties of quantum systems—for example, predicting whether an unknown quantum process satisfies time-reversal symmetry. Their tests show that quantum computers can maintain their advantages in solving certain problems, even when errors specific to quantum computers are taken into account.

The work of Huang *et al*. intertwines the ability to characterize quantum systems (*4*–*6*) with machine learning, with implications for near-term quantum computers and perhaps even quantum sensing. The introduced generalization of classical machine learning to allow quantum data as inputs allows for certain benefits; namely, difficult proofs of advantages of quantum computers become easier. However, because of the hardware required to transfer quantum data in its unperturbed state from an experiment into the quantum computer, this method may be difficult to implement in certain settings, such as the high-energy physics experiments at the Large Hadron Collider. In smaller-scale experiments, however, transduction may be reasonable—for example, in quantum-optical experiments with nitrogen-vacancy centers in diamonds (*7*), which are often designed with transporting quantum information in mind. In a related vein, this work also opens a frontier for quantum sensing that involves quantum states and may lead to better advantages (*8*). Huang *et al*. proved in detail that for the data-driven prediction of properties of quantum experiments, no classical computer will ever pose a challenge to quantum ones—and that quantum computers may soon help expand human knowledge into new echelons. ■

REFERENCES AND NOTES

1. H.-Y. Huang *et al*., *Science* **376**, 1182 (2022).
2. H. Y. Huang *et al*., *Nat. Commun.* **12**, 2631 (2021).
3. C. H. Bennett *et al*., *Phys. Rev. A* **59**, 1070 (1999).
4. S. Aaronson, *Proc. R. Soc. London A* **463**, 3089 (2007).
5. S. Aaronson, *STOC 2018: Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing* 10.1145/3188745.3188802 (2018).
6. S. Chen, J. Cotler, H.-Y. Huang, J. Li, *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, 2022), pp. 574–585.
7. M. Ruf *et al*., *J. Appl. Phys.* **130**, 070901 (2021).
8. C. L. Degen, F. Reinhard, P. Cappellaro, *Rev. Mod. Phys.* **89**, 035002 (2017).

10.1126/science.abp9885

QUANTUM COMPUTATION

Solving a puzzle with atomic qubits

A quantum computer makes light work of the maximum independent set problem

By **Monika Schleier-Smith**

Imagine that you are asked to color a map of the world. Starting with your favorite color, you endeavor to fill in as many countries as possible without giving any neighboring countries the same color. This puzzle, despite its straightforward premise, is notorious for its computational complexity. On page 1209 of this issue, Ebadi *et al*. (*1*) report a quantum algorithm for solving the puzzle—known as the maximum independent set (MIS) problem—using individual atoms trapped in optical tweezers to represent the countries on the map. The demonstration is an important milestone in the broad effort to understand which computational problems stand to benefit from quantum computers.

To date, only a few quantum algorithms have been proven to offer clear advantages over classical computers. Moreover, even in cases where quantum computers theoretically provide a benefit—such as for factoring large numbers—practical applications will require major advances in quantum hardware beyond the current state of the art. By contrast, the coloring puzzle presented by Ebadi *et al*. belongs to a large class of optimization problems (*2*) that are potentially easier to solve using near-term quantum devices (*3*) but for which the attainable quantum speedup remains largely an open question (*4*–*6*). Such optimization problems, with technological relevance in areas such as supply chain logistics, can generically be framed as minimizing what is known as a cost function. The solution can be calculated by tasking the quantum computer to minimize the energy of a system of interacting particles or qubits, where the specific problem is encoded in the structure of the interactions.

To generate the structure of interactions required to represent MIS problems, Ebadi

Department of Physics, Stanford University, Stanford, CA, USA. Email: schleier@stanford.edu

By harnessing the power of the Google Sycamore processor (pictured here), Huang *et al*. showcase the exponential advantages offered by quantum computers for analyzing data from quantum experiments.

mere classical information. The quantum data can be used by the quantum computer to predict future results while requiring far fewer experiments.

One cannot input quantum data into a classical computer. Intuitively, one may think that this would give the quantum device a straightforward advantage. But the actual scenario is more nuanced. For a classical setup, the quantum state of an experiment can be measured and used as input, with each measurement freely chosen by the classical learning algorithm. Because the classical computer can arbitrarily choose when to measure each of the experiments, then at least in principle, all the information encoded in quantum states can be accessible. For a quantum setup, the quantum computer provides a minimal but key additional capacity: a small quantum memory that enables joint measurements on two copies of quantum data.

So in both cases, all the quantum data are converted to classical information before calculation, but in slightly different manners. Joint measurements—used in the quantum-enhanced scenario—unravel correlated properties of two separate quantum systems. This fundamentally exploits quantum entanglement (when the quantum states of two or more objects are intertwined with each other) and cannot be substituted by pairs of individual measurements. Since the early days of quantum information theory, it has been known that joint measurements can help distinguish quantum states, even when the states are uncorrelated (*3*). But until recently, it was not clear just how large an advantage this exploit can give quantum computers over their classical counterparts.

Building from the research line on so-called shadow tomography (*4*–*6*), Huang *et al*. argued that joint measurements lead to substantial advantages for learning about quantum systems. Namely, the quantum-enhanced strategy is exponentially more economical in terms of the number of quantum experiments needed for predicting the outcomes of just two measurements (*6*). The authors demonstrated the advantages of a quantum learning experiment using the Google Sycamore processor. The natural scenario of quantum data learning involves a "transducer" that transports the quantum state of results from an experiment into the quantum computer. Their experiment was simulated in the same quantum processor that analyzes the data, in a lab-on-a-chip setting. Once the quantum state is prepared, it is analyzed with classical and quantum-enhanced methods.

For the optimal predictions, the exact joint measurements may be known, at least in idealized settings. However, in the real experiment, the state preparation is imperfect, as is the measurement performed. To counteract this, the quantum processing is supplemented with classical machine learning to extract the strongest signals in the presence of experimental errors. This classical-quantum hybrid approach demonstrates advantages in our capacity to learn various fundamental properties of quantum systems—for example, predicting whether an unknown quantum process satisfies time-reversal symmetry. Their tests show that quantum computers can maintain their advantages in solving certain problems, even when errors specific to quantum computers are taken into account.

The work of Huang *et al*. intertwines the ability to characterize quantum systems (*4*–*6*) with machine learning, with implications for near-term quantum computers and perhaps even quantum sensing. The introduced generalization of classical machine learning to allow quantum data as inputs allows for certain benefits; namely, difficult proofs of advantages of quantum computers become easier. However, because of the hardware required to transfer quantum data in its unperturbed state from an experiment into the quantum computer, this method may be difficult to implement in certain settings, such as the high-energy physics experiments at the Large Hadron Collider. In smaller-scale experiments, however, transduction may be reasonable—for example, in quantum-optical experiments with nitrogen-vacancy centers in diamonds (*7*), which are often designed with transporting quantum information in mind. In a related vein, this work also opens a frontier for quantum sensing that involves quantum states and may lead to better advantages (*8*). Huang *et al*. proved in detail that for the data-driven prediction of properties of quantum experiments, no classical computer will ever pose a challenge to quantum ones—and that quantum computers may soon help expand human knowledge into new echelons. ■

REFERENCES AND NOTES

1. H.-Y. Huang *et al*., *Science* **376**, 1182 (2022).
2. H. Y. Huang *et al*., *Nat. Commun.* **12**, 2631 (2021).
3. C. H. Bennett *et al*., *Phys. Rev. A* **59**, 1070 (1999).
4. S. Aaronson, *Proc. R. Soc. London A* **463**, 3089 (2007).
5. S. Aaronson, *STOC 2018: Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing* 10.1145/3188745.3188802 (2018).
6. S. Chen, J. Cotler, H.-Y. Huang, J. Li, *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, 2022), pp. 574–585.
7. M. Ruf *et al*., *J. Appl. Phys.* **130**, 070901 (2021).
8. C. L. Degen, F. Reinhard, P. Cappellaro, *Rev. Mod. Phys.* **89**, 035002 (2017).

10.1126/science.abp9885

QUANTUM COMPUTATION

Solving a puzzle with atomic qubits

A quantum computer makes light work of the maximum independent set problem

By **Monika Schleier-Smith**

Imagine that you are asked to color a map of the world. Starting with your favorite color, you endeavor to fill in as many countries as possible without giving any neighboring countries the same color. This puzzle, despite its straightforward premise, is notorious for its computational complexity. On page 1209 of this issue, Ebadi *et al*. (*1*) report a quantum algorithm for solving the puzzle—known as the maximum independent set (MIS) problem—using individual atoms trapped in optical tweezers to represent the countries on the map. The demonstration is an important milestone in the broad effort to understand which computational problems stand to benefit from quantum computers.

To date, only a few quantum algorithms have been proven to offer clear advantages over classical computers. Moreover, even in cases where quantum computers theoretically provide a benefit—such as for factoring large numbers—practical applications will require major advances in quantum hardware beyond the current state of the art. By contrast, the coloring puzzle presented by Ebadi *et al*. belongs to a large class of optimization problems (*2*) that are potentially easier to solve using near-term quantum devices (*3*) but for which the attainable quantum speedup remains largely an open question (*4*–*6*). Such optimization problems, with technological relevance in areas such as supply chain logistics, can generically be framed as minimizing what is known as a cost function. The solution can be calculated by tasking the quantum computer to minimize the energy of a system of interacting particles or qubits, where the specific problem is encoded in the structure of the interactions.

To generate the structure of interactions required to represent MIS problems, Ebadi

Department of Physics, Stanford University, Stanford, CA, USA. Email: schleier@stanford.edu

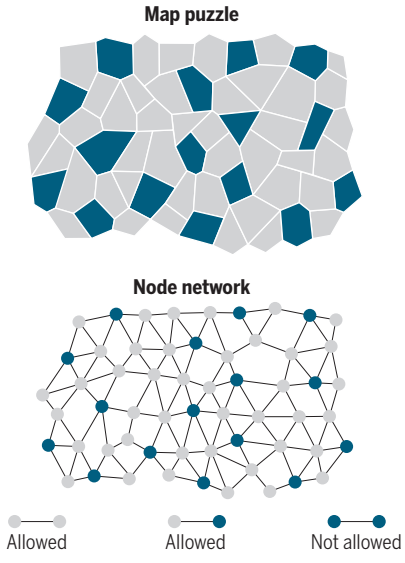
et al. used qubits encoded in the internal states of optically trapped atoms. Each atom can either be in the electronic ground state or a highly excited state known as a Rydberg state, where the electron cloud is thousands of times larger than in the ground state. An atom can be excited to the Rydberg state by a laser. However, any attempt to excite multiple neighboring atoms is constrained by strong interactions between atoms in the Rydberg state. Specifically, no more than a single atom can be excited within a minimum distance known as the blockade radius (*7*–*9*). Thus, atoms that are closer to one another than the blockade radius are equivalent to countries that share a border, where only one but not both can be colored blue, or in this case, be excited to the Rydberg state.

The method was put to test using a quantum processor with up to 289 atomic qubits, with each qubit trapped at the focus of a laser beam. By controlling the positions of the atoms, Ebadi *et al*. programmed specific instances of the MIS problem, each of which can be visualized as a graph with an atom at each node and with bonds between blockaded pairs (see the figure). They sought to solve the problem using an approach known as adiabatic quantum computation (*4*, *10*). Here, the system parameters are ramped from an initial state in which the minimum-energy configuration is simple and known to a final state where the minimum-energy configuration provides a solution for the MIS problem. In the laser-driven atomic system, depending on whether the photon energy is lower or higher than the energy of a Rydberg excitation, the minimum-energy configuration can either have all atoms in the ground state or have as many atoms in the Rydberg state as possible without violating the blockade constraint. Thus, by ramping the frequency and intensity of the lasers, the atoms are driven from their initial ground states into a configuration of Rydberg excitations that, ideally, forms an MIS.

The key to this method is the maintenance of adiabaticity within the system—to ensure that the quantum system remains in its lowest-energy state throughout the ramping process. As an analogy, think of a waiter delivering an ice cream float to a diner. If the waiter moves too quickly, the drink may spill out, yet if he moves too slowly, the ice cream will melt—both of which are undesirable from the perspective of the diner. Similarly, in the quantum experiment, the system parameters must increase slowly enough for the atoms to settle into the MIS and yet fast enough for the quantum system to maintain its coherence, which is ultimately limited by the lifetime of the Rydberg states.

Coloring a map with a quantum computer

How many blue regions can this map have without any of them sharing a border? Instead of examining all the possibilities classically, Ebadi *et al*. solved the puzzle by using a quantum computer, composed of individual atoms that can only be excited (shown as ●) if all neighboring atoms are in their ground states (●). The map puzzle is encoded as a network of nodes, which represent the regions, and connections, which represent the shared borders.



A crucial question is whether this quantum algorithm provides a speedup over classical approaches. State-of-the-art classical algorithms employ a strategy known as simulated annealing, which mimics a physical process of preparing the interacting system at a high temperature and gradually reducing the temperature to reach the lowest-energy state. In practice, neither the quantum nor the classical algorithm always succeeds in finding the optimal solution. Thus, a figure of merit for the performance of the algorithm is the average number of times (t) that the algorithm would need to be run to succeed in finding the MIS. For the classical approach, this number of iterations t_{SA} is proportional to the ratio of the number of near-perfect solutions to the number of perfect solutions, where a near-perfect solution is defined as having one fewer "country" in its set than a perfect solution.

By contrast, the performance of the quantum algorithm not only depended on how many near-perfect solutions there are for every perfect one, it also depended on the gap in energy between the lowest-energy state and the first excited state. The smaller this gap, the slower a ramp one theoretically expects to require for the system to remain in its lowest-energy state to reach the perfect solutions. In cases where a sufficiently slow ramp could be performed within the time scale of the experiment, the quantum algorithm provided a speedup compared with the classical one. Specifically, the number of tries required by the quantum algorithm to solve the MIS problem scaled as the three-fifths power of the number of tries t_{SA} required classically, meaning that if the MIS problem is made more difficult such that the time required to solve it classically increases, for example, by a factor of 32, then the time required by the quantum computer will only increase by a factor of 8.

Although the experiment by Ebadi *et al*. is not the first to explore quantum optimization algorithms (*11*–*13*), it stands out for operating both with a large number of qubits and with sufficiently coherent interactions for quantum information to spread across the entire system within the time scale of the experiment (*14*). This combination appears to be crucial for the observed quantum speedup.

An important question for future work is whether the improved scaling of the quantum algorithm persists as the difficulty of the problems is increased, for example, by increasing the number of qubits. A potential challenge is that the gap in energy separating perfect from near-perfect solutions is expected to shrink as the number of qubits grows (*3*, *11*), placing ever more stringent demands on how slowly the system parameters must be swept, and hence also on the coherence time of the experiment. One hope is to adopt the approach of an experienced waiter who moves so quickly that the drink begins to slosh, but ultimately executes just the right motions to bring it back to rest (*15*). Finding the right motions in a complex quantum system is bound to be a challenge, offering fertile ground for future research. ■

REFERENCES AND NOTES

1. S. Ebadi *et al*., *Science* **376**, 1209 (2022).
2. R. M. Karp, in *Complexity of Computer Computations*, the IBM Research Symposia Series, R. E. Miller, J. W. Thatcher, J. D. Bohlinger, Eds. (Springer, 1972), pp. 85–103.
3. A. Lucas, *Front. Phys. (Lausanne)* **2**, 5 (2014).
4. T. Albash, D. A. Lidar, *Rev. Mod. Phys*. **90**, 015002 (2018).
5. S. Bravyi, A. Kliesch, R. Koenig, E. Tang, *Phys. Rev. Lett*. **125**, 260505 (2020).
6. T. F. Rønnow *et al*., *Science* **345**, 420 (2014).
7. E. Urban *et al*., *Nat. Phys*. **5**, 110 (2009).
8. T. Wilk, A. Gaëtan *et al*., *Phys. Rev. Lett*. **104**, 010502 (2010).
9. P. Schauß *et al*., *Nature* **491**, 87 (2012).
10. E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, D. Preda, *Science* **292**, 472 (2001).
11. S. Boixo *et al*., *Nat. Phys*. **4**, 2067 (2013).
12. G. Pagano *et al*., *Proc. Natl. Acad. Sci. U.S.A*. **117**, 25396 (2020).
13. M. P. Harrigan *et al*., *Nat. Phys*. **17**, 332 (2021).
14. A. Omran *et al*., *Science* **365**, 570 (2019).
15. E. J. Crosson, D. A. Lidar, *Nat. Rev. Phys*. **3**, 466 (2021).

10.1126/science.abq3754

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

CHEMISTRY

Phosphorus through the looking glass

A key building block enables a general synthesis of chiral phosphorus drugs

By **Xavier Verdaguer**[1,2]

In the book *Alice's Adventures in Wonderland*, Alice suspected that the milk on the other side of the looking glass might not be good for the cat to drink. Perhaps unbeknownst to the writer Lewis Carroll, this subplot reflects an important aspect of chemistry—that the mirror image of a molecule is not necessarily the same as the original because of a feature known as chirality. Just like our left and right hands, the mirrored image of a chiral molecule is not superimposable to the original. On page 1230 of this issue, Forbes and Jacobsen (*1*) describe an approach for the selective synthesis of a single mirror image of chiral pharmaceutical compounds, the chirality of which is originated at a phosphorus atom. The approach may make the synthesis of highly desired chiral drugs much more attainable.

Chirality is essential to life. Amino acids and sugars are chiral and are found as single right-handed or left-handed isomers. Therefore, proteins and most biomolecules are intrinsically chiral. Chirality in molecules most often originates from carbon atoms with four different substituents in a tetrahedral disposition, with the central atom labeled as "stereogenic" or as a "stereocenter." After chemists Jean-Baptiste Biot and Louis Pasteur discovered and rationalized the phenomenon of chirality in organic compounds, others soon realized that elements beyond carbon, like phosphorus, could also be stereogenic, that is, P-stereogenic. In 1911, chemists Jakob Meisenheimer and Leo Lichtenstadt were the first to separate a P-stereogenic compound into its individual mirror isomers (*2*). Although this work initially seemed like nothing more than a scientific curiosity, it was a few decades later when P-stereogenic phosphorus ligands found applications in the industrial synthesis of levodopa (L-dopa), a compound for treating Parkinson's disease (*3*). William S. Knowles received the 2001 Nobel Prize in Chemistry for this discovery.

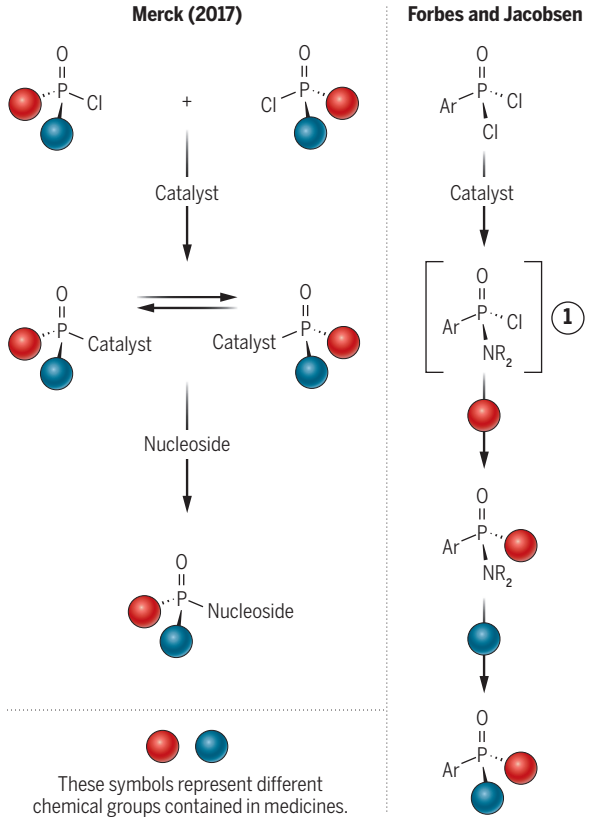
Nowadays, P-stereogenic compounds have found their way into numerous drugs. To make antitumoral and antiviral treatments more effective, nucleoside drugs are often introduced into the body as prodrugs, which are medications that become active only after entering the patient, and are used to help improve a medication's effectiveness. These phosphorus prodrugs contain a P-stereogenic atom, as in tenofovir alafenamide (*4*). Cyclic dinucleotides like cyclic guanosine monophosphate–adenosine monophosphate (cGAMP), which are promising candidates for cancer treatments, contain two P-stereogenic phosphorus atoms (*5*). The synthesis of such phosphorus derivatives in a precise and stereocontrolled manner is of vital importance because it will determine the performance of the final drug.

Until very recently, the synthesis of single-handed P-stereogenic compounds relied on the use of other chiral molecules known as stoichiometric chiral auxiliaries. These are initially attached to the phosphorus atom but must be discarded by the end of the synthesis (*6*, *7*). An initial breakthrough in this field came in 2017 when chemists at Merck developed a highly specific catalyst for the synthesis of a P-stereogenic prodrug compound using a dynamic kinetic resolution strategy (*8*). Here, a mixture of equal parts left-handed and right-handed starting material was used. The catalyst attaches to the starting material and exerts two functions: It provides a bias for the reaction with the nucleoside such that the right isomer reacts faster than the left one, and it interconverts the two mirror isomers, allowing the nonreacted left isomer to be converted to the desired right-handed product. Very recently, a similar strategy was used in the P(III) phosphoramidite coupling of oligonucleotides using a chiral phosphoric acid catalyst. The approach allowed the stereocontrolled synthesis of cyclic dinucleotides like cGAMP (*9*). However, both methods are substrate specific and do not serve as a general strategy for the synthesis of P-stereogenic compounds.

Forbes and Jacobsen describe an alternative and general approach based on a symmetrical—that is, nonchiral—starting material that is modified to introduce chirality by using a so-called desymmetrization reaction (see the figure). The authors used a starting material

Single-handed phosphorus drugs

Merck's strategy uses a 50/50 mixture of right- and left-handed starting material. This method is highly substrate specific. Forbes and Jacobsen use a nonchiral starting material. The key single-handed building block (**1**) allows for the synthesis of multiple chiral phosphorus drugs through sequential substitution.



GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

[1]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology (BIST), 08028 Barcelona, Spain. [2]Departament de Química Inorgànica i Orgànica, Secció Química Orgànica, Universitat de Barcelona, 08028 Barcelona, Spain. Email: xavier.verdaguer@irbbarcelona.org

with two chlorides. The catalyst, a ureasulfinamide, promoted the substitution of only one of the chlorides by an amine to yield a highly enriched single-handed P-stereogenic chlorophosphonamide. This is a highly versatile P-stereogenic building block because the chloride group and the amino group have the orthogonal reactivity necessary for the stereocontrolled sequential substitution of the two groups. Basic reagents react with inversion of configuration with the chloride, whereas the amine can be substituted under acidic conditions for alcohols. This approach allows the synthesis of a large variety of P-stereogenic compounds, including phosphonates, phosphinates, phosphonamidates, and phosphonate thioesters. The potential of this methodology is demonstrated by Forbes and Jacobsen in the synthesis of a few P-stereogenic drugs.

Phosphine chloride compounds are highly reactive. To harness the reactivity of dichlorophosphinyl derivatives in a desymmetrization reaction is an impressive achievement. However, this success comes with a caveat. The versatile chlorophosphinamide building block cannot be isolated in pure form because it is prone to racemization, that is, an equilibration to a 50/50 mixture of right- and left-handed isomers. This hampers the enrichment of the isomer

"This approach allows the synthesis of a large variety of P-stereogenic compounds..."

of interest and ultimately might limit the use of this method. Future developments in the catalytic synthesis of P-stereogenic compounds should seek methods with improved selectivity and stable P-stereogenic intermediates that can compete with state-of-the-art traditional chiral auxiliary approaches. The work of Forbes and Jacobsen will encourage chemists to work on further developments and to explore the wonders that await us beyond the looking glass. ■

REFERENCES AND NOTES

1. K. C. Forbes, E. N. Jacobsen, *Science* **376**, 1230 (2022).
2. J. Meisenheimer, L. Lichtenstadt, *Ber. Dtsch. Chem. Ges.* **44**, 356 (1911).
3. W. S. Knowles, *Acc. Chem. Res.* **16**, 106 (1983).
4. C. McGuigan *et al.*, *J. Med. Chem.* **48**, 3504 (2005).
5. H. Zhang, Q.-D. You, X.-L. Xu, *J. Med. Chem.* **63**, 3785 (2020).
6. K. W. Knouse *et al.*, *Science* **361**, 1234 (2018).
7. S. Lemouzy, L. Giordano, D. Hérault, G. Buono, *Eur. J. Org. Chem.* **2020**, 3351 (2020).
8. D. A. DiRocco *et al.*, *Science* **356**, 426 (2017).
9. A. L. Featherston *et al.*, *Science* **371**, 702 (2021).

10.1126/science.abq5073

MOLECULAR BIOLOGY

Solving the nuclear pore puzzle

Using a battery of tools, the architecture of the nuclear pore complex is revealed

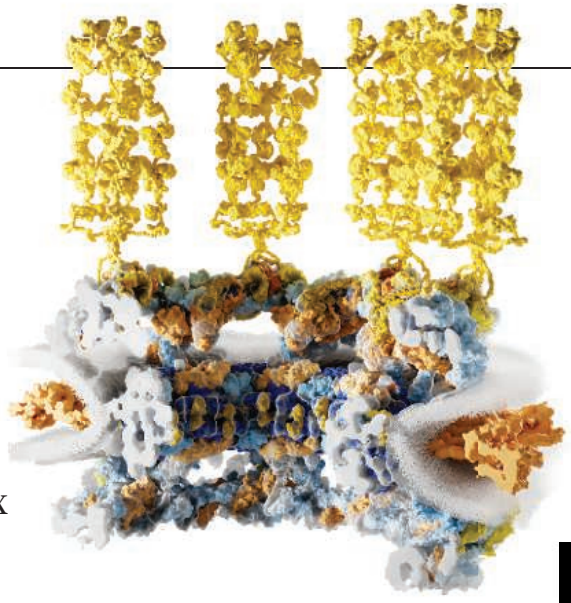


ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*; PDB DATA: ANDRÉ HOELZ AND MARTIN BECK

By **Thomas U. Schwartz**

In eukaryotic cells, the genome is sequestered in the nucleus, shielded from the cytoplasm by the double-layered nuclear envelope (NE). Transport of macromolecules across the NE occurs through nuclear pore complexes (NPCs), which perforate the NE at ~200 to 2000 positions (*1*–*3*). Ions and molecules up to ~40 kDa diffuse through NPCs, whereas larger cargo selectively associate with soluble nuclear transport factors to be ferried through the central NPC channel (*4*). But it has been unclear how NPCs exactly control the transport of a vast array of different substrates, including soluble proteins, embedded membrane proteins, RNAs, and even some viral capsids. On pages 1174, 1175, 1176, 1177, and 1178 of this issue, Bley *et al.* (*5*), Petrovic *et al.* (*6*), Mosalaganti *et al.* (*7*), Zhu *et al.* (*8*), and Fontana *et al.* (*9*), respectively, now provide molecular structures, in unprecedented detail, of how NPCs are built. These findings will enable approaches to further dissect the many NPC functions.

With an estimated size of 60 to 120 MDa, depending on the eukaryotic species, NPCs are elaborate protein assemblies. The 600 to 1000 individual proteins, collectively called nucleoporins (NUPs), are organized around a central eightfold rotational symmetry into protomers (also sometimes referred to as "spokes"). Protomer segments build four concentric rings, referencing their position relative to the pore opening in the NE: cytoplasmic ring (CR), inner ring (IR), nucleoplasmic ring (NR), and luminal ring (LR). Further, each protomer consists of subcomplexes, which are defined as biochemically stable entities into which NPCs disassemble during cell division and NE breakdown. About half of the core scaffold, which encompasses ~25 different NUPs, has previously been positioned within the NPC structure, primarily through a combination of x-ray structural analysis of individual subcomplexes and cryo–electron tomography (cryo-ET) reconstructions (or maps) of the entire NPC (*1*–*3*). The papers in this issue now fill many of the remaining gaps.

Starting from an established protocol for the cryo-ET analysis of human NPCs, Mosalaganti *et al.* improved the resolution of the CR and IR from ~2.3 to ~1.2 nm. In addition, the authors used artificial intelligence–based structure prediction and subcomplex modeling to interpret the improved cryo-ET map. The NPC scaffold also serves as the anchor for phenylalanine-glycine (FG) repeats that extend their fibrillar extensions into the central channel of the NPC to form the principal transport barrier (*10*). Mosalaganti *et al.* can now position, in addition to the central NUP62-NUP58-NUP54 complex, the second main FG-containing assembly—the CR-attached NUP62-NUP88-NUP214 complex—completing the anchor points for the FG network.

The same cryo-ET map was also interpreted by Petrovic *et al.* and Bley *et al.*; these authors used extensive experimental data to support the fitting of experimental structures. Petrovic *et al.* focus on linker NUPs that connect subcomplexes within the NPC. They establish how the large, stacked helical proteins NUP188 and NUP205 interact competitively with NUP93. Together with a number of weaker interaction motifs, the authors provide a linker map of the IR.

Bley *et al.* focus on the CR and specifically on the attachment of so-called cytoplasmic filaments, which are particularly important for mRNA export. NUP358 is an extended, multidomain protein, and five copies anchor to the CR through the amino-terminal 75 kDa α-solenoid element, solved by Bley *et al.* by using x-ray crystallography. The flexibly linked NUP93-NUP205 pair, hitherto considered a component of the IR, is also

Massachusetts Institute of Technology, Department of Biology, Cambridge, MA, USA. Email: tus@mit.edu

with two chlorides. The catalyst, a ureasulfinamide, promoted the substitution of only one of the chlorides by an amine to yield a highly enriched single-handed P-stereogenic chlorophosphonamide. This is a highly versatile P-stereogenic building block because the chloride group and the amino group have the orthogonal reactivity necessary for the stereocontrolled sequential substitution of the two groups. Basic reagents react with inversion of configuration with the chloride, whereas the amine can be substituted under acidic conditions for alcohols. This approach allows the synthesis of a large variety of P-stereogenic compounds, including phosphonates, phosphinates, phosphonamidates, and phosphonate thioesters. The potential of this methodology is demonstrated by Forbes and Jacobsen in the synthesis of a few P-stereogenic drugs.

Phosphine chloride compounds are highly reactive. To harness the reactivity of dichlorophosphinyl derivatives in a desymmetrization reaction is an impressive achievement. However, this success comes with a caveat. The versatile chlorophosphinamide building block cannot be isolated in pure form because it is prone to racemization, that is, an equilibration to a 50/50 mixture of right- and left-handed isomers. This hampers the enrichment of the isomer

"This approach allows the synthesis of a large variety of P-stereogenic compounds..."

of interest and ultimately might limit the use of this method. Future developments in the catalytic synthesis of P-stereogenic compounds should seek methods with improved selectivity and stable P-stereogenic intermediates that can compete with state-of-the-art traditional chiral auxiliary approaches. The work of Forbes and Jacobsen will encourage chemists to work on further developments and to explore the wonders that await us beyond the looking glass. ■

REFERENCES AND NOTES

1. K. C. Forbes, E. N. Jacobsen, *Science* **376**, 1230 (2022).
2. J. Meisenheimer, L. Lichtenstadt, *Ber. Dtsch. Chem. Ges.* **44**, 356 (1911).
3. W. S. Knowles, *Acc. Chem. Res.* **16**, 106 (1983).
4. C. McGuigan *et al.*, *J. Med. Chem.* **48**, 3504 (2005).
5. H. Zhang, Q.-D. You, X.-L. Xu, *J. Med. Chem.* **63**, 3785 (2020).
6. K. W. Knouse *et al.*, *Science* **361**, 1234 (2018).
7. S. Lemouzy, L. Giordano, D. Hérault, G. Buono, *Eur. J. Org. Chem.* **2020**, 3351 (2020).
8. D. A. DiRocco *et al.*, *Science* **356**, 426 (2017).
9. A. L. Featherston *et al.*, *Science* **371**, 702 (2021).

10.1126/science.abq5073

MOLECULAR BIOLOGY

Solving the nuclear pore puzzle

Using a battery of tools, the architecture of the nuclear pore complex is revealed

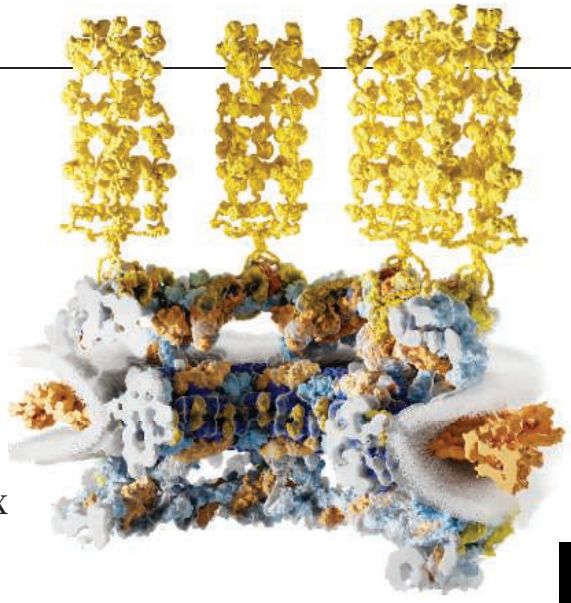


ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*; PDB DATA: ANDRÉ HOELZ AND MARTIN BECK

By **Thomas U. Schwartz**

In eukaryotic cells, the genome is sequestered in the nucleus, shielded from the cytoplasm by the double-layered nuclear envelope (NE). Transport of macromolecules across the NE occurs through nuclear pore complexes (NPCs), which perforate the NE at ~200 to 2000 positions (*1*–*3*). Ions and molecules up to ~40 kDa diffuse through NPCs, whereas larger cargo selectively associate with soluble nuclear transport factors to be ferried through the central NPC channel (*4*). But it has been unclear how NPCs exactly control the transport of a vast array of different substrates, including soluble proteins, embedded membrane proteins, RNAs, and even some viral capsids. On pages 1174, 1175, 1176, 1177, and 1178 of this issue, Bley *et al.* (*5*), Petrovic *et al.* (*6*), Mosalaganti *et al.* (*7*), Zhu *et al.* (*8*), and Fontana *et al.* (*9*), respectively, now provide molecular structures, in unprecedented detail, of how NPCs are built. These findings will enable approaches to further dissect the many NPC functions.

With an estimated size of 60 to 120 MDa, depending on the eukaryotic species, NPCs are elaborate protein assemblies. The 600 to 1000 individual proteins, collectively called nucleoporins (NUPs), are organized around a central eightfold rotational symmetry into protomers (also sometimes referred to as "spokes"). Protomer segments build four concentric rings, referencing their position relative to the pore opening in the NE: cytoplasmic ring (CR), inner ring (IR), nucleoplasmic ring (NR), and luminal ring (LR). Further, each protomer consists of subcomplexes, which are defined as biochemically stable entities into which NPCs disassemble during cell division and NE breakdown. About half of the core scaffold, which encompasses ~25 different NUPs, has previously been positioned within the NPC structure, primarily through a combination of x-ray structural analysis of individual subcomplexes and cryo–electron tomography (cryo-ET) reconstructions (or maps) of the entire NPC (*1*–*3*). The papers in this issue now fill many of the remaining gaps.

Starting from an established protocol for the cryo-ET analysis of human NPCs, Mosalaganti *et al.* improved the resolution of the CR and IR from ~2.3 to ~1.2 nm. In addition, the authors used artificial intelligence–based structure prediction and subcomplex modeling to interpret the improved cryo-ET map. The NPC scaffold also serves as the anchor for phenylalanine-glycine (FG) repeats that extend their fibrillar extensions into the central channel of the NPC to form the principal transport barrier (*10*). Mosalaganti *et al.* can now position, in addition to the central NUP62-NUP58-NUP54 complex, the second main FG-containing assembly—the CR-attached NUP62-NUP88-NUP214 complex—completing the anchor points for the FG network.

The same cryo-ET map was also interpreted by Petrovic *et al.* and Bley *et al.*; these authors used extensive experimental data to support the fitting of experimental structures. Petrovic *et al.* focus on linker NUPs that connect subcomplexes within the NPC. They establish how the large, stacked helical proteins NUP188 and NUP205 interact competitively with NUP93. Together with a number of weaker interaction motifs, the authors provide a linker map of the IR.

Bley *et al.* focus on the CR and specifically on the attachment of so-called cytoplasmic filaments, which are particularly important for mRNA export. NUP358 is an extended, multidomain protein, and five copies anchor to the CR through the amino-terminal 75 kDa α-solenoid element, solved by Bley *et al.* by using x-ray crystallography. The flexibly linked NUP93-NUP205 pair, hitherto considered a component of the IR, is also

Massachusetts Institute of Technology, Department of Biology, Cambridge, MA, USA. Email: tus@mit.edu

Newly resolved components of the human nuclear pore complex include the luminal ring (orange), cytoplasmic filaments (yellow), and lipid membrane (white). The structure known in 2016 is in blue.

positioned in two locations of the CR and one position of the NR, in entirely different environments than in the IR. One rationale for this moonlighting function, at least in the CR, is the role of NUP93 in anchoring the cytoplasmic NUP62-NUP88-NUP214 FG-containing complex, which is analogous to the central NUP62-NUP58-NUP54 FG-containing complex in the IR. Although only one NUP62-NUP88-NUP214 complex is modestly resolved in the protomer cryo-ET map, stochiometric and steric arguments suggest that there are two copies per protomer (or 16 copies per CR). Overall, the three studies of human NPCs arrive at a reassuringly similar conclusion about the positioning of NUPs.

Zhu *et al.* and Fontana *et al.* pursued a different approach. Oocytes from the African clawed frog (*Xenopus laevis*) have nuclear membranes packed with NPCs, which enables single-particle reconstruction by use of cryo–electron microscopy (cryo-EM). In cryo-EM, many more particles are averaged compared with that in cryo-ET, which boosts resolution. The studies of Zhu *et al.* and Fontana *et al.* visualize a CR protomer at up to ~4 and ~7 Å, respectively. With well-resolved secondary structure elements, and even residue-level resolution for substantial parts of the protomer, these studies enable detailed analysis. The findings suggest that the core scaffold of the vertebrate NPC is well conserved. As previously determined, the nine-membered core Y complex forms two concentric, interwoven eight-membered rings (*1*–*3*). The five NUP358 amino-terminal α-solenoid structures per protomer are very similar in both analyses, as are two NUP205 molecules. Beyond that, the structures differ. Fontana *et al.* resolve two NUP62-NUP88-NUP214 complexes, as predicted by the cryo-ET–based studies, but Zhu *et al.* do not. Instead, they exclusively position two NUP93 molecules attached to the CR. These differences may be explained by different data processing strategies, slightly different sample preparations, or both. More likely, though, they point to a larger issue, which is the intrinsic heterogeneity of NPCs.

A limitation of these studies is that to achieve high resolution, many NPCs, and specifically individual protomers, need to be averaged. Given the high degree of structural redundancy among scaffold NUPs, it is expected that NPCs are conformationally and constitutionally heterogenous (*3*, *11*). A recent structural analysis of NPCs from the yeast *Saccharomyces cerevisiae* provided the clearest example of this to date, in which different compositional classes of NPCs were visualized (*12*). The more ambiguous parts of the different reconstructions published in this issue may point to inadvertent averaging over nonidentical protomers, resulting in partially distorted densities. This is difficult to separate from flexibility, which commonly reduces map contrast.

A prominent example of conformational heterogeneity within the human NPC is the recent discovery that the central channel is substantially wider in intact cells when compared with partially purified NPCs (57 versus 43 nm, respectively) (*13*, *14*). This conformational change is largely confined to the IR and LR, not the two outer rings. It may thus not be a coincidence that the CR is the best-resolved NPC element, likely reflecting higher homogeneity. How can the IR adopt such different conformations? The consensus from the studies of Mosalaganti *et al.* and Petrovic *et al.* is that the protomers do not structurally change but rather move as rigid blocks, enabled by highly flexible linkers that keep the protomers connected. This way, peripheral channels are established that enable transport of embedded membrane proteins from the outer to inner nuclear membrane, and past the NPC.

Overall, the studies of Bley *et al.*, Petrovic *et al.*, Mosalaganti *et al.*, Zhu *et al.*, and Fontana *et al.* substantially advance the knowledge of the assembly of NPCs, primarily in vertebrates. NPCs can now be functionally and structurally probed in unprecedented detail. Although much of the approach by the different researchers can be likened to solving a giant jigsaw puzzle, it is surprising that several identical pieces fit into many different positions. It will be interesting to tease apart the potential functional relevance of this unusual binding behavior and to reveal NPC biology at the level of detail comparable with that of other central problems in cell biology. ■

REFERENCES AND NOTES

1. B. Hampoelz *et al.*, *Annu. Rev. Biophys.* **48**, 515 (2019).
2. D. H. Lin, A. Hoelz, *Annu. Rev. Biochem.* **88**, 725 (2019).
3. J. Fernandez-Martinez, M. P. Rout, *Trends Biochem. Sci.* **46**, 595 (2021).
4. C. E. Wing, H. Y. J. Fung, Y. M. Chook, *Nat. Rev. Mol. Cell Biol.* **23**, 307 (2022).
5. C. J. Bley, A. Hoelz, *Science* **376**, eabm9129 (2022).
6. S. Petrovic, A. Hoelz, *Science* **376**, eabm9798 (2022).
7. S. Mosalaganti, M. Beck, *Science* **376**, eabm9506 (2022).
8. X. Zhu, Y. Shi, *Science* **376**, eabl8280 (2022).
9. P. Fontana, H. Wu, *Science* **376**, eabm9326 (2022).
10. H. B. Schmidt, D. Görlich, *Trends Biochem. Sci.* **41**, 46 (2016).
11. K. E. Knockenhauer, T. U. Schwartz, *Cell* **164**, 1162 (2016).
12. C. W. Akey *et al.*, *Cell* **185**, 361 (2022).
13. A. P. Schuller *et al.*, *Nature* **598**, 667 (2021).
14. C. E. Zimmerli *et al.*, *Science* **374**, eabd9776 (2021).

ACKNOWLEDGMENTS

T.U.S. is supported by National Institutes of Health grant R35-GM141834.

10.1126/science.abq4792

METABOLISM

Aligning mealtimes to live longer

Calorie restriction, fasting, and circadian rhythms sync together for a long, healthy life in mice

By **Shaunak Deota** *and* **Satchidananda Panda**

Calorie restriction (CR) involves chronic reduction of energy intake by 20 to 40% without inducing malnutrition (*1*). CR extends life span in multiple animal models and reduces the risk of age-associated disorders, most of which arise from metabolic dysfunction and inflammation. However, extended daily fasting or aligning daily meal timing to the active period, even without reducing energy intake, can also improve health and increase life span in model organisms. On page 1192 of this issue, Acosta-Rodríguez *et al.* (*2*) reveal the specific contribution of fasting and timing of calorie-reduced meals to the efficacy of CR, as estimated by life-span extension in male mice.

In most rodent CR studies, the control group is fed ad libitum (AL), whereas the CR animals are fed a single meal per day that contains ~20 to 40% fewer calories than the AL group consumes. The CR animals eat most of their daily ration within 2 hours (*3*). Thus, CR studies inadvertently introduce timed or time-restricted feeding and prolonged daily fasting, both of which can improve health and delay aging independently of CR (*4*). Disentangling the effect of CR, fasting, and time of feeding on rodent life span is not easy. Acosta-Rodríguez *et al.* developed a system that automatically delivers a specific quantity of food in a bolus, as in standard CR studies, or in small meals at specific times (*3*). All mice were housed under 12 hours of light and dark to synchronize circadian rhythms (internal body clock), and all cages were equipped with wheels that measured voluntary wheel-running.

To assess the impact of calories alone on life span, they split the CR ration (30% reduced calories relative to the AL group) into

The Salk Institute for Biological Studies, La Jolla, CA, USA. Email: satchin@salk.edu

Newly resolved components of the human nuclear pore complex include the luminal ring (orange), cytoplasmic filaments (yellow), and lipid membrane (white). The structure known in 2016 is in blue.

positioned in two locations of the CR and one position of the NR, in entirely different environments than in the IR. One rationale for this moonlighting function, at least in the CR, is the role of NUP93 in anchoring the cytoplasmic NUP62-NUP88-NUP214 FG-containing complex, which is analogous to the central NUP62-NUP58-NUP54 FG-containing complex in the IR. Although only one NUP62-NUP88-NUP214 complex is modestly resolved in the protomer cryo-ET map, stochiometric and steric arguments suggest that there are two copies per protomer (or 16 copies per CR). Overall, the three studies of human NPCs arrive at a reassuringly similar conclusion about the positioning of NUPs.

Zhu *et al.* and Fontana *et al.* pursued a different approach. Oocytes from the African clawed frog (*Xenopus laevis*) have nuclear membranes packed with NPCs, which enables single-particle reconstruction by use of cryo–electron microscopy (cryo-EM). In cryo-EM, many more particles are averaged compared with that in cryo-ET, which boosts resolution. The studies of Zhu *et al.* and Fontana *et al.* visualize a CR protomer at up to ~4 and ~7 Å, respectively. With well-resolved secondary structure elements, and even residue-level resolution for substantial parts of the protomer, these studies enable detailed analysis. The findings suggest that the core scaffold of the vertebrate NPC is well conserved. As previously determined, the nine-membered core Y complex forms two concentric, interwoven eight-membered rings (*1*–*3*). The five NUP358 amino-terminal α-solenoid structures per protomer are very similar in both analyses, as are two NUP205 molecules. Beyond that, the structures differ. Fontana *et al.* resolve two NUP62-NUP88-NUP214 complexes, as predicted by the cryo-ET–based studies, but Zhu *et al.* do not. Instead, they exclusively position two NUP93 molecules attached to the CR. These differences may be explained by different data processing strategies, slightly different sample preparations, or both. More likely, though, they point to a larger issue, which is the intrinsic heterogeneity of NPCs.

A limitation of these studies is that to achieve high resolution, many NPCs, and specifically individual protomers, need to be averaged. Given the high degree of structural redundancy among scaffold NUPs, it is expected that NPCs are conformationally and constitutionally heterogenous (*3*, *11*). A recent structural analysis of NPCs from the yeast *Saccharomyces cerevisiae* provided the clearest example of this to date, in which different compositional classes of NPCs were visualized (*12*). The more ambiguous parts of the different reconstructions published in this issue may point to inadvertent averaging over nonidentical protomers, resulting in partially distorted densities. This is difficult to separate from flexibility, which commonly reduces map contrast.

A prominent example of conformational heterogeneity within the human NPC is the recent discovery that the central channel is substantially wider in intact cells when compared with partially purified NPCs (57 versus 43 nm, respectively) (*13*, *14*). This conformational change is largely confined to the IR and LR, not the two outer rings. It may thus not be a coincidence that the CR is the best-resolved NPC element, likely reflecting higher homogeneity. How can the IR adopt such different conformations? The consensus from the studies of Mosalaganti *et al.* and Petrovic *et al.* is that the protomers do not structurally change but rather move as rigid blocks, enabled by highly flexible linkers that keep the protomers connected. This way, peripheral channels are established that enable transport of embedded membrane proteins from the outer to inner nuclear membrane, and past the NPC.

Overall, the studies of Bley *et al.*, Petrovic *et al.*, Mosalaganti *et al.*, Zhu *et al.*, and Fontana *et al.* substantially advance the knowledge of the assembly of NPCs, primarily in vertebrates. NPCs can now be functionally and structurally probed in unprecedented detail. Although much of the approach by the different researchers can be likened to solving a giant jigsaw puzzle, it is surprising that several identical pieces fit into many different positions. It will be interesting to tease apart the potential functional relevance of this unusual binding behavior and to reveal NPC biology at the level of detail comparable with that of other central problems in cell biology. ■

REFERENCES AND NOTES

1. B. Hampoelz *et al.*, *Annu. Rev. Biophys.* **48**, 515 (2019).
2. D. H. Lin, A. Hoelz, *Annu. Rev. Biochem.* **88**, 725 (2019).
3. J. Fernandez-Martinez, M. P. Rout, *Trends Biochem. Sci.* **46**, 595 (2021).
4. C. E. Wing, H. Y. J. Fung, Y. M. Chook, *Nat. Rev. Mol. Cell Biol.* **23**, 307 (2022).
5. C. J. Bley, A. Hoelz, *Science* **376**, eabm9129 (2022).
6. S. Petrovic, A. Hoelz, *Science* **376**, eabm9798 (2022).
7. S. Mosalaganti, M. Beck, *Science* **376**, eabm9506 (2022).
8. X. Zhu, Y. Shi, *Science* **376**, eabl8280 (2022).
9. P. Fontana, H. Wu, *Science* **376**, eabm9326 (2022).
10. H. B. Schmidt, D. Görlich, *Trends Biochem. Sci.* **41**, 46 (2016).
11. K. E. Knockenhauer, T. U. Schwartz, *Cell* **164**, 1162 (2016).
12. C. W. Akey *et al.*, *Cell* **185**, 361 (2022).
13. A. P. Schuller *et al.*, *Nature* **598**, 667 (2021).
14. C. E. Zimmerli *et al.*, *Science* **374**, eabd9776 (2021).

ACKNOWLEDGMENTS

T.U.S. is supported by National Institutes of Health grant R35-GM141834.

10.1126/science.abq4792

METABOLISM

Aligning mealtimes to live longer

Calorie restriction, fasting, and circadian rhythms sync together for a long, healthy life in mice

By **Shaunak Deota** *and* **Satchidananda Panda**

Calorie restriction (CR) involves chronic reduction of energy intake by 20 to 40% without inducing malnutrition (*1*). CR extends life span in multiple animal models and reduces the risk of age-associated disorders, most of which arise from metabolic dysfunction and inflammation. However, extended daily fasting or aligning daily meal timing to the active period, even without reducing energy intake, can also improve health and increase life span in model organisms. On page 1192 of this issue, Acosta-Rodríguez *et al.* (*2*) reveal the specific contribution of fasting and timing of calorie-reduced meals to the efficacy of CR, as estimated by life-span extension in male mice.

In most rodent CR studies, the control group is fed ad libitum (AL), whereas the CR animals are fed a single meal per day that contains ~20 to 40% fewer calories than the AL group consumes. The CR animals eat most of their daily ration within 2 hours (*3*). Thus, CR studies inadvertently introduce timed or time-restricted feeding and prolonged daily fasting, both of which can improve health and delay aging independently of CR (*4*). Disentangling the effect of CR, fasting, and time of feeding on rodent life span is not easy. Acosta-Rodríguez *et al.* developed a system that automatically delivers a specific quantity of food in a bolus, as in standard CR studies, or in small meals at specific times (*3*). All mice were housed under 12 hours of light and dark to synchronize circadian rhythms (internal body clock), and all cages were equipped with wheels that measured voluntary wheel-running.

To assess the impact of calories alone on life span, they split the CR ration (30% reduced calories relative to the AL group) into

The Salk Institute for Biological Studies, La Jolla, CA, USA. Email: satchin@salk.edu

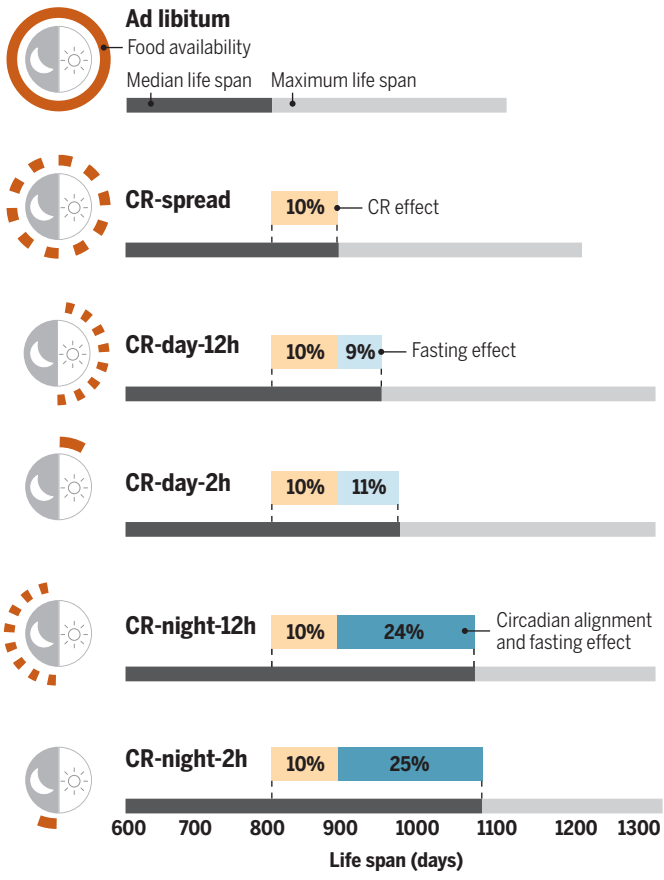
nine equal meals delivered every 160 min (CR-spread), thus minimizing the chance of prolonged fasting. To assess the contribution of the length of fasting, they delivered the same CR ration within a 2-hour window (CR-2h or 22-hour fasting) or split into eight equal meals delivered every 90 min over 12 hours (CR-12h or 12-hour fasting). To test the impact of the CR ration fed during the day or night, they delivered the CR-2h and CR-12h in the day (CR-day-2h, CR-day-12h) or night (CR-night-2h, CR-night-12h) (see the figure).

Despite all CR groups receiving the same quality and quantity of diet, their life spans differed. In the CR-spread group, the median life span was 10% longer than that of the AL group. All other CR groups that had 12-hour or 22-hour fasting lived longer than the CR-spread group, demonstrating that fasting can boost the life-extending effect of CR. However, the median life span of CR-day-2h or CR-day-12h mice fed during their daytime rest period was extended by 20%, whereas mice in the CR-night-2h and CR-night-12h groups (when they are active) lived almost 35% longer than mice in the AL group (*2*). Thus, the life-extending effect of fasting is further boosted when it overlaps with the circadian sleep period.

Despite eating less, all mice at middle age in the CR groups were more active than mice in the AL group. CR and fasting increase the production of ketone bodies from the liver, which can act on the circadian clock in the brain to increase food-seeking activity (*5*). Physical activity and exercise exert pleiotropic health benefits in mice and humans (*6*). Accordingly, as the mice aged, more-active mice also lived longer. However, among the CR groups, the CR-spread mice were more active in older age and yet lived the shortest amount of time. The CR-spread mice had slightly more daytime activity, coinciding with their mealtime, and there was a trend toward daytime activity and reduced life span. The CR-day mice also had slightly more daytime activity, and they did not live as long as the CR-night mice. Although sleep was not measured, being more active during the day likely disrupts sleep in nocturnal rodents. Altogether, these results imply that maintaining a robust sleep-wake and fasting-feeding cycle aligned with the circadian clock, can boost the life span–extending effect of CR.

Meal times and calorie restriction affect longevity

Circadian alignment of feeding to the most active part of the day (nighttime in mice) and fasting boosts the life span–extending effects of calorie restriction (CR).



The fasting-feeding cycle in all CR groups except CR-spread likely sustained better circadian rhythms. Although the light-dark cycle entrains the central circadian clock and supports sleep (*7*), the fasting-feeding cycle sustains robust circadian rhythms in peripheral organs (*8*). Hence, optimum alignment of fasting and feeding with the light-dark cycle can maintain robust circadian rhythms, which in turn activates numerous pathways in different organs at the optimum time (*9*).

Aligning fasting-feeding with the circadian clock–programmed sleep-wake cycle to gain health benefits has some precedence. In day-active fruit flies, nightly fasting leads to life-span extension compared with daytime fasting (*10*). Mutations in circadian clock genes disrupt feeding and sleeping patterns and also dampen CR-dependent life-span extension in fruit flies and mice (*11*, *12*). Even when mice are fed an isocaloric diet, those fed during the active phase have improved health and extended life span (*8*).

To investigate the cause of death, Acosta-Rodríguez *et al.* examined mice that were found dead of old age or moribund mice that were euthanized. They found that cancer was the major cause of death in all groups and that liver cancer was the most prevalent type, suggesting that CR delayed the onset or severity of cancer. To gain clues about the underlying mechanism, the authors probed the liver transcriptomes of all cohorts. CR attenuated the age-associated gene expression changes observed in AL mice and maintained a gene expression signature indicative of better metabolic homeostasis. The CR-night group predominantly exhibited changes in immune gene expression, suggesting reduced inflammation. This group also had more genes expressed in alignment with circadian rhythms compared with CR-day groups. Circadian gene expression temporally optimizes cellular processes, which may partly explain the longer life span of CR-night groups.

Beyond the liver transcriptomes, there was little clue about life-span effects in different CR groups. Body weight and body composition were comparable among all CR groups. Likewise, in humans, 25% CR with or without 8-hour time-restricted eating (8-hour CR or 16-hour fast) for 1 year led to similar weight loss (*13*). Conversely, in weight-stabilized humans, 6-hour time-restricted eating improved cardiometabolic health compared with habitual eating within 12-hours (*14*). This implies that changes in body weight or composition in response to dietary interventions that involve CR, fasting, or meal timing may not accurately predict the overall effectiveness of these interventions, although perhaps trials with larger cohorts and more restrictive timing are needed. Studies of plasma and tissues from people undergoing CR, fasting, or meal-time restriction hold untapped potential for understanding the system-wide impact of these dietary interventions and predicting the extent to which they can prevent or manage chronic diseases. ■

REFERENCES AND NOTES

1. L. K. Heilbronn, E. Ravussin, *Am. J. Clin. Nutr.* **78**, 361 (2003).
2. V. Acosta-Rodríguez *et al.*, *Science* **376**, 1192 (2022).
3. V. A. Acosta-Rodríguez *et al.*, *Cell Metab.* **26**, 267 (2017).
4. V. D. Longo, S. Panda, *Cell Metab.* **23**, 1048 (2016).
5. R. Chavan *et al.*, *Nat. Commun.* **7**, 10580 (2016).
6. J. A. Sanford *et al.*, *Cell* **181**, 1464 (2020).
7. M. Hatori, S. Panda, *Trends Mol. Med.* **16**, 435 (2010).
8. E. N. C. Manoogian *et al.*, *Endocr. Rev.* **43**, 405 (2022).
9. L. S. Mure *et al.*, *Science* **359**, eaao0318 (2018).
10. M. Ulgherait *et al.*, *Nature* **598**, 353 (2021).
11. S. D. Katewa *et al.*, *Cell Metab.* **23**, 143 (2016).
12. S. A. Patel *et al.*, *FASEB J.* **30**, 1634 (2016).
13. D. Liu *et al.*, *N. Engl. J. Med.* **386**, 1495 (2022).
14. E. F. Sutton *et al.*, *Cell Metab.* **27**, 1212 (2018).

10.1126/science.adc8824

GRAPHIC: N. DESAI/*SCIENCE*

VIEWPOINT: COVID-19

Immuno-epidemiology and the predictability of viral evolution

Understanding viral evolution depends on a synthesis of evolutionary biology and immuno-epidemiology

By **Chadi M. Saad-Roy[1], C. Jessica E. Metcalf[2,3], Bryan T. Grenfell[2,3]**

Despite much recent progress in modeling the epidemiology and evolution of acute viruses, a full quantitative synthesis of viral eco-evolutionary dynamics remains elusive. The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has stimulated vast research efforts into measuring viral dynamics and genetics, across scales from individual hosts to global circulation. In parallel, understanding determinants of individual immune protection against infection and severe disease has been a major research focus. However, the interaction between population immunity and viral dynamics has been much less studied—a crucial gap because this interaction will strongly influence SARS-CoV-2 evolution over time. Clarifying the biology of transmission at different scales, and in particular the impact of immunity on transmission, will define the epidemiological context (current spread and population risk), as well as the epidemiological and evolutionary implications, of immune escape.

To probe likely evolutionary trajectories, it is necessary to understand how immunity intersects with transmission and thus population outcomes, especially in partially immune individuals. Building such a "phylodynamic" synthesis (*1*) requires a revolution in cross-scale understanding of how individual immune kinetics translate to immuno-epidemiology. The resulting frameworks will provide a more mechanistic basis for exploring the predictability of viral evolutionary dynamics. Such mechanistic approaches synthesize findings across disciplines, and they will provide a principled way to integrate information across scales and inform data analyses.

Phylodynamic models meld the epidemiological and evolutionary dynamics of pathogens, often with underlying host immune kinetics. Such frameworks have been applied across a range of acute and chronic pathogens (*1*, *2*), and often focus on influenza eco-evolutionary dynamics as a model for acute partially immunizing infections. In particular, the immune escape of seasonal variants of human influenza A virus has been used to study the impact of population immunity on viral population dynamics. The simplest conceptual qualitative phylodynamic models for immune escape (*1*) posited that spread of escape variants would be maximized at intermediate immune pressure, through a trade-off between transmission and immune selection. That is, if there is no immune pressure, then viral abundance may be high but selection for immune escape is absent. Conversely, if there is strong immune pressure, then selection may be high but viral abundance very low; this would also limit adaptive evolution for immune escape.

More detailed dynamic models for influenza have addressed quantitative phylodynamic interactions across scales (from within hosts to global spread). These models also explored what limits viral diversity to the observed phylogeny that emerges from antigenic drift (changes in surface proteins that lead to immune escape), given the huge variation generated by error-prone viral reproduction (*3*, *4*). An array of related framings of immunodynamics in various guises [such as strain-transcending immunity (*3*) and repeated selective sweeps driven by herd immunity (*4*)] have been explored as candidates to explain these phylodynamic patterns for influenza A virus.

Phylodynamic models have been widely applied beyond seasonal influenza (*2*). In particular, the dynamics of seasonal human coronaviruses (HCoVs) are becoming increasingly salient, especially owing to the continued spread of SARS-CoV-2. For example, the 229E HCoV exhibits antigenic drift (*5*), which could have implications for the emergence of SARS-CoV-2 variants. Indeed, the ongoing COVID-19 pandemic underlines the major impact not only of immune escape evolution, but also of rapid selection for (surprisingly large) increases in viral transmission rate. Although the acquisition of human-to-human transmission by influenza viruses that caused pandemics is arguably an example of a selection on transmission, this is not as evident for seasonal influenza.

A key issue for next-generation phylodynamics is to determine how selection at different levels (within-host, transmission chains, population-level, globally) translates into population outcomes, and how it is modulated by host immunity. There are important processes present at each biological scale, from the emergence of variants to their global spread (supplementary fig. S1). Within individual hosts, variants arise through mutation and/or recombination (*1*). The ability of variants to replicate (which is necessary for successful transmission) will be affected by their cellular tropism and by the efficiency with which they can enter cells and transmit across tissues. For example, increases in angiotensin-converting enzyme 2 (ACE2) avidity may lead to enhanced SARS-CoV-2 transmissibility (*6*) and may also alter tissue tropism, e.g., for the upper respiratory tract (*7*). Tropism could also depend on host immune responses; adaptive immunity (acquired through infection or vaccination) may also shape viral load trajectories [e.g., for influenza (*8*)] and lead to selection for immune-escape variants [e.g., the Omicron variant (*9*)]. Tropism and adaptive immunity also likely play a role in the clinical severity of infections (and reinfections) with variants.

The combination of individual immune phenotypes and their impact on viral shedding affects the transmission of variants and determines whether they can cause breakthrough infections in vaccinated individuals. Simultaneously, antigenic drift or waning immunity may influence susceptibility to (re)infection. In turn, these lead to fitness advantages for variants with either increased immune escape or transmissibility. Individual characteristics of immunity might be particularly important in shaping the features of the virus. For example, selection of influenza viruses within immunologically competent hosts is less strong owing to asynchrony between viral growth and immune response (*8*), whereas prolonged carriage in immunocompromised hosts could result in variants (*10*); this may also be the case for SARS-CoV-2 (*11*).

At the population level, the presence of many immune individuals could limit viral spread through indirect protection. To prevent the establishment of variants with increased transmissibility (but with little immune escape) once they have emerged, a higher proportion of immune individuals are required than that needed to limit transmission of the original virus. Indirect

[1]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA. [2]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. [3]Princeton School of Public and International Affairs, Princeton University, Princeton, NJ, USA.
Email: csaadroy@princeton.edu; grenfell@princeton.edu

protection from immune escape variants will depend on specific host and viral properties, including susceptibility to reinfection and transmission from breakthrough infection. Additionally, global movement seeds new SARS-CoV-2 outbreaks, potentially in regions with vastly different levels of preexisting infection-induced or vaccinal immunity and other epidemiological factors. Depending upon this immune landscape, population-level selection may favor variants with either higher transmissibility or increased immune escape. Inequitable vaccine coverage increases global infection levels, which increases the likelihood for variants with either of these characteristics.

To untangle the effects of immunity on transmission, it will be crucial to quantify the cross-scale impact of host immune responses on epidemiology and viral evolution. Although several studies have made progress in this general direction (*8*, *10*), there remain important gaps that need to be addressed. In particular, epidemiological studies that simultaneously quantify immunity and transmission across populations and at different biological scales are necessary, along with the development of cross-scale modeling approaches. There are a number of study designs and associated measurements that could address these gaps and determine the impact of immunity on cross-scale phylodynamics (see the figure). Long-term longitudinal studies that monitor immunity across cohorts and ages, in addition to measuring viral loads, sequence variation, and household transmission, are essential. Using these data in cross-scale modeling frameworks could elucidate the impact of immunity on transmission and viral variation.

The phylodynamics of the COVID-19 pandemic have provided many—often nasty—surprises compared to expectations honed by the dynamics of other viruses, notably influenza virus. For example, the observation that some hosts sicken with COVID-19, but do not then strongly seroconvert (*12*), complicates interpretation of measuring antibody titers for dissecting the impact of population immunity. Assuming these immunologically "cryptic" individuals are capable of virus transmission, do they contribute to net evolution of the virus (e.g., owing to the presence of unmeasured immunity)? It is also unclear how host immune responses besides neutralizing antibodies (such as T cell immunity, which is harder to measure) affect evolutionary trajectories. Moreover, whether existing population immunity to endemic viruses creates a cross-protective firewall against emergent pathogen spillover, and how this might occur, needs further investigation (as does the potential impact of animal reservoirs).

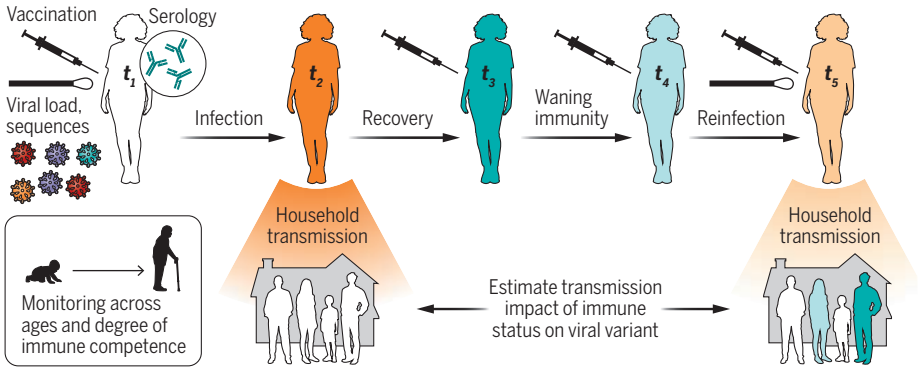
Perhaps the central question for future SARS-CoV-2 variants is the relationship between transmission rate, immune escape, and clinical severity. Specifically, is there a "worst case" viral genotype that combines high transmission, immune escape, and severity? This will depend in a complex way on—still partially understood—cross-scale viral and immune dynamics. Unfortunately, the virus is likely to explore these phylodynamics before they can be predicted, especially given continued transmission arising from global inequities in vaccine supply.

One general strategy for exploring these complexities is to generate phylodynamic case law by broadening studies to explore the phylodynamics of multiple pathogens using recently developed multiplex immunological methods (*13*) with viral sequencing. Analyzing immunity against multiple pathogens and variants will also address issues such as antigenic imprinting (the immunity conferred after recovery from the first infection) in influenza, the potential contribution of immunologically cryptic individuals to pathogen evolution, serotype interactions in dengue, and the polymicrobial impact of nonpharmaceutical interventions used to mitigate COVID-19.

New phylodynamic models and data structures should be developed to answer these questions. For example, studies have pioneered statistical methods to determine transmission patterns and evolutionary history from sequence data (*14*), and a fitness-based model to predict future incidence of influenza clades (*15*). Combining either approach with models that account for within-host kinetics and immunoepidemiology may be a particularly fruitful avenue to unravel the impact of host immunity on pathogen eco-evolutionary dynamics. For SARS-CoV-2 specifically, it is important to quantify immune landscapes at all spatial scales (local, regional, and global). Crucially, this quantification will enable the development of tools and infrastructure that will aid in predicting future epidemic and evolutionary trajectories. These suggested approaches echo proposals for a Global Immunological Observatory (i.e., regular mass blood sampling to monitor immunological signatures) (*13*), but with an important evolutionary focus to elucidate the cross-scale impacts of immunity on pathogen dynamics. More generally, such immuno-epidemiology and viral discovery could help elucidate the phylodynamics and vaccinology of a much wider range of pathogens. ■

Studying immuno-epidemiology

Cohort studies are needed to understand the impact of immunity on phylodynamic evolutionary patterns. There are two classes of unknowns: the impact of vaccination for individuals with different degrees of immunity, and the impact of vaccination on subsequent susceptibility to and transmissibility of breakthrough infections.



GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

REFERENCES AND NOTES

1. B. T. Grenfell *et al.*, *Science* **303**, 327 (2004).
2. E. M. Volz, K. Koelle, T. Bedford, *PLOS Comput. Biol.* **9**, e1002947 (2013).
3. N. M. Ferguson, A. P. Galvani, R. M. Bush, *Nature* **422**, 428 (2003).
4. K. Koelle *et al.*, *Science* **314**, 1898 (2006).
5. R. T. Eguia *et al.*, *PLOS Pathog.* **17**, e1009453 (2021).
6. J. Zahradník *et al.*, *Nat. Microbiol.* **6**, 1188 (2021).
7. K. P. Y. Hui *et al.*, *Nature* **603**, 715 (2022).
8. D. H. Morris *et al.*, *eLife* **9**, e62105 (2020).
9. D. Planas *et al.*, *Nature* **602**, 671 (2022).
10. K. S. Xue *et al.*, *eLife* **6**, e26875 (2017).
11. L. Corey *et al.*, *N. Engl. J. Med.* **385**, 562 (2021).
12. W. Liu *et al.*, *Emerg. Infect. Dis.* **27**, 2454 (2021).
13. M. J. Mina *et al.*, *eLife* **9**, e58989 (2020).
14. R. A. Smith, E. L. Ionides, A. A. King, *Mol. Biol. Evol.* **34**, 2065 (2017).
15. M. Łuksza, M. Lässig, *Nature* **507**, 57 (2014).

ACKNOWLEDGMENTS

The authors thank A. Spaulding, D. Douek, and A. McDermott for discussions. The authors are funded by the Natural Sciences and Engineering Research Council of Canada through a Postgraduate Doctoral Scholarship (C.M.S.-R.); a Charlotte Elizabeth Procter Fellowship of Princeton University (C.M.S.-R); the US Centers for Disease Control and Prevention (B.T.G.); and the Flu Lab (B.T.G.).

10.1126/science.abn9410

POLICY FORUM

CLIMATE CHANGE

Land management can contribute to net zero

The voluntary carbon market needs to embrace changes for the land sector

By **Ruth DeFries [1,2], Richie Ahuja[3], Julio Friedman[4,5], Doria R. Gordon[3,6], Steven P. Hamburg[3], Suzi Kerr[3], James Mwangi[7,8], Carlijn Nouwen[7], Nitin Pandit[9]**

Demand for credits on the voluntary carbon market is poised to surge as corporations implement net-zero commitments. Approximately half of all credits issued from 2000 to 2021 on the voluntary carbon market related to land use, mostly from forest projects (fig. S1) (*1*). Credits from the land sector pose challenges for accounting and for meeting multiple criteria. We propose three pathways to overcome shortcomings in the carbon market, improve integrity of credits, and promote long-lasting change to achieve nontrivial climate mitigation and co-benefits from the land sector: (i) target major sources of land-based emissions by increasing activities that reduce or avoid non-CO_2 greenhouse gas (GHG) emissions; (ii) promote longevity of low-GHG land management by ensuring that locally relevant co-benefits accrue to local land users; and (iii) encourage region-wide over individual project-based activities to promote systemic change, provide equitable access to benefits, enable realistic accounting, and scale opportunities for emissions reductions.

Agriculture, forestry, and land use account for ~20% of global anthropogenic GHGs (from enteric fermentation in livestock and manure, agricultural soils, crop burning, deforestation, cropland degradation, and rice cultivation in decreasing order of emissions) (*2*). Reducing emissions from this sector is essential to address the goals of the Paris Agreement. In addition, restoration and improved management of multiple types of ecosystems can provide opportunities to remove carbon from the atmosphere and store it in the biosphere, although the magnitudes are uncertain and potential land conflicts raise serious concerns. Estimates indicate that up to 60% of emissions from agriculture relative to 2030 business-as-usual projections and 110% of emissions from the forestry sector are technologically and economically feasible to reduce (*3*).

Multiple approaches can economically incentivize reduced emissions and carbon sequestration from land management, including fines for violating regulations, subsidies and tax credits, capped emission in cap-and-trade programs, and payments. All of these approaches can play a role in a transition to low-GHG land management. The current net-zero commitments and explosive growth in the voluntary carbon market offer an unprecedented opportunity to bring finance to the land sector for mitigating climate change, if criteria and accounting can overcome problems of low integrity and inequitable benefit sharing.

"A shortcoming of the current project-based carbon market is its small-scale, piecemeal approach through individual projects."

The integrity of land-based credits in the voluntary carbon market has been historically uneven. Buyer confidence can only occur with high-integrity credits—in other words, assurance that credits represent documented, actual reductions, avoidance, or removal of GHGs that would otherwise not occur. This integrity needs to improve substantially for these types of investments to effectively reduce GHG emissions or remove them from the atmosphere in voluntary carbon markets.

Multiple factors contribute to low confidence in carbon market credits, including difficulties with assurances of additionality (whether interventions to reduce emissions or store carbon would occur in the absence of revenue from the carbon market), permanence (the integrity of how reversed benefits of avoided or stored emissions are addressed), leakage (whether reductions in one place are displaced by emissions to another place), and quantification (whether reductions are accurately quantified relative to an appropriate baseline and reported). Another factor contributing to low confidence is the potential for displacement and inequitable benefit sharing with marginalized and Indigenous peoples in places with conflicts between customary and statutory land tenure, asymmetric power relations, or poor governance. Such negative outcomes have occurred with previous implementation of financing for forest-based climate mitigation (*4*).

In the absence of improved integrity, carbon markets are likely to shift toward projects that buyers perceive as higher quality, such as long-term geologic storage from carbon capture and storage, that more easily meet the criteria. Without alternative pathways to improve integrity of credits from the land sector, opportunities for the sector to contribute to cost-effective climate mitigation and its potential co-benefits for biodiversity and local livelihoods could be negated.

NON-CO_2 GHGS

The Sixth Assessment of the Intergovernmental Panel on Climate Change highlights the opportunity to reduce non-CO_2 GHGs, particularly methane (CH_4), as a mechanism to reduce the rate of warming over the near to medium term and provide relief from climate change. Land-based activities are substantial anthropogenic sources of non-CO_2 GHGs, which drive the bulk of warming between now and 2050. Enteric fermentation, manure, and rice cultivation together emit more CH_4 than fossil fuels, and agriculture emits roughly four times as much nitrous oxide (N_2O) as fossil fuels (table S1). Until the recent pledge to reduce CH_4 emissions from a diversity of sectors by 30% by 2030, adopted by more than a hundred countries at the 26th Conference of Parties in Glasgow, there has been no systematic effort to drive down these critical non-CO_2 emissions.

Currently, credits issued and number of projects in the voluntary carbon market are overwhelmingly related to projects that avoid or reduce CO_2 emissions or sequester

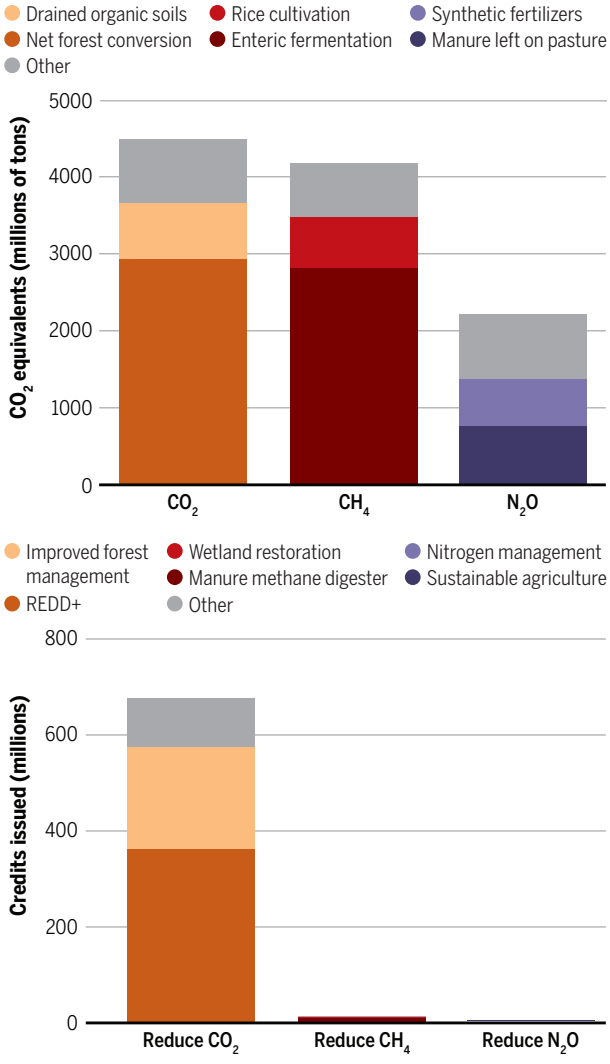
[1]Department of Ecology, Evolution, and Environmental Biology, Columbia University, New York, NY, USA. [2]Climate School, Columbia University, New York, NY, USA. [3]Environmental Defense Fund, New York, NY, USA. [4]Center for Global Energy Policy, Columbia University, New York, NY, USA. [5]Carbon Direct LLC, New York, NY, USA. [6]Department of Biology, University of Florida, Gainsville, FL, USA. [7]Climate Action Platform Africa, Nairobi, Kenya. [8]Dalberg Group, Nairobi, Kenya. [9]Independent consultant, Vienna, VA, USA. Email: rd2402@columbia.edu

carbon (see the figure and tables S2 and S3). Land use–related CH_4 and N_2O cause more than half of the climate impacts from the land sector. However, only 2% of the credits issued for land-based projects from 1996 to 2021 aimed to reduce emissions of CH_4, and practically none addressed N_2O (*1*). In terms of number of all land-based projects, 17% aimed to reduce CH_4 and 1% N_2O (fig. S2). As an example of the imbalance between sources of emissions and credits in the market to reduce those emissions, enteric fermentation emits almost as much CO_2 equivalents (100-year time frame) globally as does forest conversion, yet only about 300 credits (two projects) addressed CH_4 reductions through feed additives in contrast to more than 300 million (140 projects) for forest-focused projects, despite the relatively low or negative costs per tonne of CO_2 equivalent of many of these activities.

For the carbon market to make substantial contributions to reducing emissions of multiple GHGs, the disproportionately high reliance on forest-related projects needs to be augmented by increased focus on enteric fermentation in livestock, irrigation management, and other agricultural projects. In particular, projects to reduce and avoid CH_4 are underrepresented in comparison to its contribution to land-based emissions. Methods to establish accurate baselines will help legitimize these initiatives.

Ample emissions, insufficient credits

(Top) Global emissions are shown from land-based sources of CO_2, CH_4, and N_2O in 2019 (in CO_2 equivalents on a 100-year time frame) (*15*). (Bottom) Credits issued from all four major registries in the voluntary carbon market from 1996 through November 2021 (*1*) are categorized by the main greenhouse gas affected by the project. The top two highest-emission sources (top) and project types for credits issued (bottom) are shown for each gas (along with "other" indicating all source and project types not encompassed by the top two categories shown for each gas). Approximately 10% of credits issued for forest projects are buffers for reserve credits in the event of reversal. Tables S2 and S3 list all land-based sources and project types. REDD+, Reducing emissions from deforestation and forest degradation.



LOCAL BENEFITS

Initiatives that reduce, avoid, or remove emissions benefit people indirectly through the global effect on climate. Local people involved in land management where the projects are implemented do not necessarily benefit directly beyond financial payments, which are often meager after accounting for transactions costs and low carbon prices. However, some buyers will pay premiums for projects with social and environmental co-benefits that contribute to Sustainable Development Goals. Co-benefits can accrue locally, such as employment opportunities. Other co-benefits, such as habitat conservation, affect local land users only indirectly or could have negative impacts such as increased human-wildlife conflict or displacement. In those cases, local land users are less likely to participate over the long term or across larger regions, resulting in unrealized potential for climate mitigation.

Evidence from land management projects indicates stronger outcomes for conservation if local land users receive tangible benefits and participate in regional land use and management decisions (*5*). Direct benefits to local land users through sustained returns of financial capital, tangible ecosystem services, or cultural value are necessary for land-based activities to successfully reduce, avoid, or remove emissions over the long term and at the scale required to meaningfully contribute to climate mitigation. Without such direct benefits, carbon mitigation efforts are likely to be short-lived and small-scale with limited effect on climate mitigation. With direct benefits, land users can experience the advantages and promote shifts to low-GHG practices within their communities.

A carbon market that facilitates transformations to low-GHG land management will value activities that are participatory and locally perceived as beneficial. For example, large-scale tree-planting schemes that aim to improve carbon sequestration and livelihoods are more effective if they include tree species preferred by local people for fuelwood, fodder, and grazing (*6*). Similarly, cookstoves that reduce the need for fuelwood, improve air quality, and save labor can shift norms toward more carbon-efficient sources of energy for cooking, and agroforestry can provide income for the local population while storing almost as much carbon as native forests (*7*). This focus on local needs and preferences will likely prove more effective for climate mitigation than short-term projects with more indirect co-benefits. Carbon market standards should ensure that local benefits occur.

Security of land tenure underpins the ability of the carbon market to both deliver benefits to local land users and create confidence for investors in land-use initiatives. Insecure tenure, gaps between de facto and de jure tenure, and colonial legacies of state control over land are common throughout the world, particularly in developing regions. In such places, which account for as much as 65% of the world's total land area (*8*), the carbon market can incentivize local land users to participate only if outside parties will not lay claim to the benefits. Robust land-based initiatives in the carbon market need to clarify and ensure that benefits flow to those who manage the land.

LARGE AREAS, REGIONALLY BASED

A shortcoming of the current project-based carbon market is its small-scale, piecemeal approach through individual projects. Consequently, the market has difficulty fostering transitions to low-GHG land management over larger areas for a meaningful aggregated impact.

GRAPHIC: K. FRANKLIN/*SCIENCE*

Jurisdictional and regional approaches, which aim to align governments, businesses, nongovernmental organizations, and local stakeholders around common goals for land management within an administrative unit, are becoming more common and offer possibilities for overcoming some of the shortcomings of a project-based approach [see (*9*) for examples]. Implementation of a jurisdictional approach faces hurdles in places with poor governance, corruption, and lack of enforcement, but these approaches potentially provide an alternative to individual projects if participants in the carbon market can overcome these hurdles.

Unlike individual project-based efforts, jurisdictions or region-wide organizations can foster low-GHG land management through credit and access to inputs for land users to alter feed for livestock, plant trees on farms, improve fertilizer management, eliminate crop residue burning, or other practices that require upfront investments. These efforts would be coupled with enforced regulations within jurisdictions against illegal deforestation and fire and other incentives to foster a transition to low-GHG land management. For example, municipality-level initiatives in the Brazilian Amazon successfully reduced deforestation by linking credit to farmers with municipality-level deforestation rates to implement the federal Plan to Prevent and Control Amazon Deforestation.

A carbon market that purchases credits from land users in jurisdictions that promote low-GHG land management could also help address the existing disincentive to potential suppliers from high transactions costs. More land users in a jurisdiction could aggregate their projects and reduce costs. Supply of high-quality carbon credits has been hampered by high costs for individual project development, monitoring, and other transactions (*10*). Small-scale farmers, which globally constitute 84% of all farmers (*11*), and Indigenous communities, which manage a quarter of the world's land area (*12*), are at a disadvantage to participate in the current carbon market, yet could make outsized contributions to effective climate mitigation. Pooled capabilities at a jurisdictional or regional scale to develop, monitor, and report could help facilitate participation of small-scale land users and promote durable transformation in land management.

Although the evidence base is only beginning to develop, jurisdictional and regional approaches potentially lower transaction costs and reduce, though do not eliminate, problems of additionality and leakage (*9*). At a project level, additionality is compromised when voluntary participation in response to an economic incentive is biased toward land users who already intended to change practices. This "adverse selection" problem diminishes as actors are grouped within larger jurisdictions and jurisdiction-wide support for low-GHG land management enables other land users to participate. Ambitious baselines below business as usual and dynamic adjustments in a structure predefined by crediting institutions could help reduce the problem of adverse selection at a jurisdictional level. Jurisdiction-wide monitoring alleviates the problem of leakage within jurisdictions because credits are issued only for net changes at the jurisdictional scale. Leakage could still occur with neighboring jurisdictions and across national boundaries, a problem that also occurs with project-based markets and which is very difficult to quantify and address through market or other mechanisms.

High-resolution remote sensing and artificial intelligence provide a new, low-cost way for jurisdictions and regions to monitor land management at a parcel level repeatedly across large areas. Many companies are pursuing means to both track land-use changes and combine that information with payment delivery platforms that land users can access through mobile phones. Efforts to pool project development, technical advice, monitoring, and payments are more feasible now than in the early days of the Clean Development Mechanism and voluntary carbon market. These tools can enable markets globally and help ensure that credits represent actual climate mitigation.

Despite these potential benefits of jurisdictional and regional approaches for changing land management norms across large areas and addressing additionality and leakage, an overriding concern for both project-based and jurisdictional approaches is the need for fair and equitable benefit sharing. Land rights and governance are deeply political and involve entrenched power dynamics (*13*). It is unrealistic to expect that the carbon market can resolve centuries-old problems of insecure land rights and procedural injustice. As a small step toward the larger goal of reducing these injustices, buyers on the carbon market can insist that strong and verified social safeguards, principles of informed consent, and fair benefit sharing are enforced as fundamental requirements for registration. These safeguards are not currently applied uniformly across the standards for certifying credits for the voluntary carbon market (*14*). Even when safeguards are applied, they are not a guarantee against unfair and inequitable practices, particularly where governments control forest lands that local people use customarily. More rigor is needed to be able to assure buyers that they are purchasing from places where land conflicts and governance are fairly and equitably managed.

Where governance structures make such efforts possible, jurisdictional and regional approaches potentially reduce transaction costs for developing baselines, monitoring, and reporting; enable wider participation from land users; shift land use norms; and enhance integrity for buyers in the voluntary carbon market. As jurisdictional approaches are rapidly evolving, the research community plays a key role in developing the evidence base about problems and successes in implementation, particularly regarding fair and equitable benefit sharing in places with insecure land rights.

The urgency of the climate crisis demands effective interventions to reduce, avoid, and remove emissions in all sectors. Low-GHG land management is not a panacea. Yet, with efforts to reduce and remove emissions in other sectors, it can contribute to a durable transformation toward addressing the crisis while providing local benefits for people and ecosystems. ■

REFERENCES AND NOTES

1. I. So, M. Elias, B. Haya, Voluntary Registry Offsets Database, University of California, Berkeley (2021); https://gspp.berkeley.edu/faculty-and-impact/centers/cepp/projects/berkeley-carbon-trading-project/offsets-database.
2. H. Ritchie, M. Roser, "CO_2 and greenhouse gas emissions," Our World in Data (2020); https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions.
3. Anon, "Pathways to a low-carbon economy: Version 2 of the global greenhouse gas abatement cost curve" (McKinsey and Company, 2009).
4. W. Carton, A. Asiyanbi, S. Beck, H. J. Buck, J. F. Lund, *Wiley Interdiscip. Rev. Clim. Change* **11**, e671 (2020).
5. J. T. Erbaugh *et al.*, *Nat. Ecol. Evol.* **4**, 1472 (2020).
6. F. Fleischman *et al.*, *Bioscience* **70**, 947 (2020).
7. D. Reang *et al.*, *J. Environ. Manage.* **298**, 113470 (2021).
8. M. Hurlbert, J. Krishnaswamy, F. X. Johnson, J. E. Rodríguez-Morales, Z. Zommers, "Risk management and decision making in relation to sustainable development" (Intergovernmental Panel on Climate Change, 2019).
9. M. von Essen, E. F. Lambin, *Front. Ecol. Environ.* **19**, 159 (2021).
10. T.-H. D. Phan, R. Brouwer, M. D. Davidson, *Ecol. Econ.* **133**, 1 (2017).
11. V. Ricciardi, Z. Mehrabi, H. Wittman, D. James, N. Ramankutty, *Nat. Sustain.* **4**, 651 (2021).
12. S. T. Garnett *et al.*, *Nat. Sustain.* **1**, 369 (2018).
13. F. Fleischman *et al.*, *Curr. Opin. Environ. Sustain.* **51**, 7 (2021).
14. D. Broekhoff, R. Spalding-Fecher, *Carbon Manag.* **12**, 635 (2021).
15. FAOSTAT, Food and Agriculture Organization, Ed.; https://www.fao.org/faostat/en/#data/GT (Rome, 2021).

ACKNOWLEDGMENTS

S.P.H., D.R.G., and S.K. were supported through a gift to Environmental Defense Fund from the Bezos Earth Fund. S. Schwartzman and R. Lubowski provided helpful comments on the manuscript. The research was conducted when N.P.'s affiliation was Ashoka Trust for Research in Ecology and Environment, Bengaluru, India.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo0613

10.1126/science.abo0613



Lindy Elkins-Tanton, principal investigator of NASA's Psyche mission, discusses small-bodies missions.

PHOTO: BILL INGALLS/NASA

SCIENCE LIVES

To Psyche and beyond

The future of US innovation will depend on teamwork, maintains a planetary scientist in a new memoir

By **Vijaysree Venkatraman**

This summer, a spacecraft will begin its 3.4-year journey toward Psyche, a huge metal-rich asteroid located between Mars and Jupiter. Most other asteroids—there are ~1,500,000 of them in the asteroid belt—are rocky or icy bodies, but Psyche is thought to be the exposed nickel-iron core of an early planet. This metallic world, 173 miles at its widest, could offer planetary scientists a glimpse into the early days of the Solar System and the formation of Earth.

In her new memoir, *A Portrait of the Scientist as a Young Woman*, Lindy Elkins-Tanton, the principal investigator of the NASA-funded Psyche mission, tells the engaging story of her life in science—from her tentative beginnings as an undergraduate student researcher in geology to her current position as a planetary scientist at Arizona State University's School of Earth and Space Exploration. The book also offers insights into the workings of academia, followed by suggestions for restructuring the research enterprise to transform the pace of innovation and education in the United States.

Elkins-Tanton was born into an upper middle-class family in Ithaca, New York, in 1956. Although she had material comforts, she did not have a happy childhood. Her relationship with her mother was fraught, and she battled depression in her youth. Quoting the poet Natalie Diaz, she writes: "You are not the sum of your injuries." "But perhaps I am the sum of the injuries that I have overcome," she adds. The realization that each of us is only a tiny part of a vast, unexplored Universe has both comforted and motivated her throughout the years.

The author writes engagingly of her own research and reveals how she and other researchers tease out truths about the Universe from bits of rock. Her investigations of the Siberian flood basalts—the result of Earth's largest land-based volcanic event, which occurred 250 million years ago—have revealed, for example, that the eruptions produced gases not unlike the ozone-depleting halocarbons humans produce today.

If we imagine our 4568-million-year-old Solar System as a 24-hour day, the upcoming Psyche mission could help us learn more about the building materials of rocky planets that were formed within the first 20 seconds. The launch of the spacecraft involves the efforts of some 800 people and 11 years of work. "While we can't go back in time, we *can* bridge that gulf of space," Elkins-Tanton writes. "We can send a robot to find out what Psyche really is."

Turning her attention to the broader business of modern research, Elkins-Tanton identifies the "hero model" of academic science as a major barrier to progress. In most universities, the leading scholar in any given area of research has ownership of a pyramid of resources dedicated to a given topic, she maintains. These "heroes" have an outsize influence on how knowledge is created, how it is funded, and how it is adopted and regulated by society. It is time to bid goodbye to these heroes, she argues, and create an organizational culture that supports teams, knowledge goals, and societal outcomes.

"The stereotype of a lone man in a lone lab making brilliant progress is a lie: every scientist needs to work with, contend with, and convince their community, and also, it's so often now a woman and not a man," she writes. To address big challenges in science and society, we need the breadth of ideas that comes from a diversity of voices.

With the research community moving toward a more multidisciplinary, team-based model, Elkins-Tanton says it is also time for a new model of university education. She compares the traditional ways of teaching science and math at the undergraduate level to "trying to train dogs by using electric collars"—lots of judgment, little encouragement. Like graduate students, undergraduates should be taught research skills such as how to ask questions, synthesize information, and render opinions, she argues.

Bullying and sexual harassment in academia are other topics that are discussed in detail. Should successful or important people in academia be allowed to persist in a community when they are harming others? Her answer is an unequivocal no. Her stand on this topic, she writes, comes partly from her own experience of childhood abuse.

This memoir chronicles the journey of one woman in science but is also a rallying cry to make academia a more supportive and diverse workplace so that the research community can better address the societal and scientific challenges of the 21st century. By the end of the book, Elkins-Tanton will have readers enthused about the upcoming Psyche mission and leave them with a greater appreciation of the teamwork that made it possible. ■



A Portrait of the Scientist as a Young Woman
Lindy Elkins-Tanton
William Morrow, 2022.
272 pp.

The reviewer is a science journalist in Boston, MA 02144, USA. Email: v.vijaysree@gmail.com

10.1126/science.abq6713

HISTORY OF MEDICINE

The surgeon and the soldiers

Plastic surgery pioneer Harold Gillies transformed facial reconstruction during World War I

By **Cathy Newman**

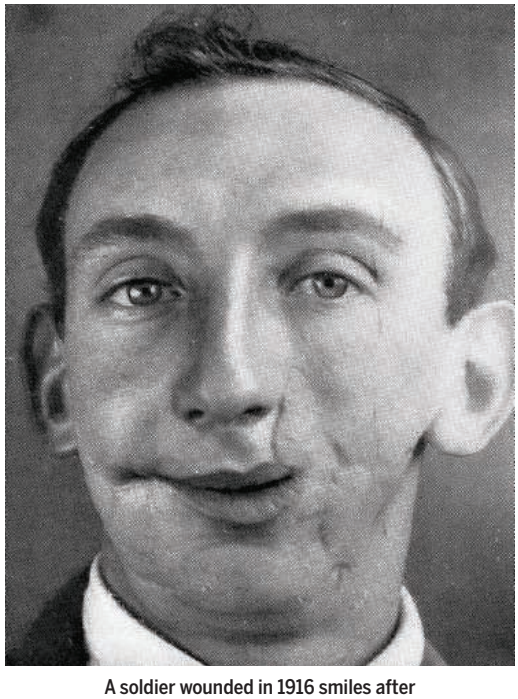
A soldier who sustained serious trauma to the face during World War I bore a soul-crushing liability. Those missing a leg or arm were apt to evoke sympathy. But those unlucky enough to "lose face"—in this instance, the term is literal as well as figurative—more often provoked disgust. Children, medical historian Lindsey Fitzharris writes in *The Facemaker*, would flee from the sight of their fathers. One wounded corporal, upon catching sight of his face in a mirror, wrote his fiancée to break their engagement. In Germany, they were called *Menschen ohne Gesicht* (men without faces); in France, *les gueules cassées* (broken faces).

Military weaponry in World War I far outstripped the ameliorations of military medicine. The deadly, disfiguring arsenal included ammunition with magnesium fuses, shrapnel, mortar bombs, caustic gases such as chlorine and phosgene, and flame throwers. Wood and canvas airplanes, meanwhile, were essentially stacks of tinder that fed fires started by exploding gas tanks. Most airmen carried a pistol to end their own lives in the event that the plane they were piloting caught fire, notes Fitzharris.

It is poignant and revealing that surgeon Harold Gillies banned mirrors in his wards. The New Zealand–born, University of Cambridge–educated pioneer in the emergent field of plastic surgery at the heart of Fitzharris's story understood the searing trauma of facial injury and disfigurement.

Before enlisting in the Royal Army Medical Corps in 1915, Gillies had worked in the posh private practice of a laryngologist on call to London's Royal Opera House, where his most memorable experience was ministering to a ballerina who had accidentally sat on a pair of scissors. "The brutal hothouse of frontline surgery" that was the Great War was different altogether. There was no textbook, no instruction manual in the radical reconstruction that many soldiers' injuries required, nor even the knowledge of how to administer anesthesia to a patient with a gaping hole in the middle of his face.

Skin grafting had been done in the 15th century to replace a nose tip sliced off by a dueling sword, but Gillies was working on a much larger canvas and understood that success depended on establishing and maintaining blood supply and warding off infection. He refined the procedure by lifting, but not severing, the graft from the donor site, stitching it into a tube, and moving the free end to the area of injury: his tubed pedicle became a reconstructive mainstay.



A soldier wounded in 1916 smiles after facial reconstruction surgery.

PHOTO: SCIENCE & SOCIETY PICTURE LIBRARY/GETTY IMAGES

But technical expertise is only part of Gillies's legacy. His creation of the Queen's Hospital at Sidcup in 1917—the first hospital dedicated to facial reconstruction—was distinguished by its multispecialty staff, which created a fertile climate for innovation and collaboration. He employed artists and photographers who could document operative technique and the before-and-after renditions of surgery, dentists conversant in jaw anatomy and physiology, caring nurses and orderlies, and even a barber trained to shave faces marked by scars and missing flesh.

The Facemaker: A Visionary Surgeon's Battle to Mend the Disfigured Soldiers of World War I
Lindsey Fitzharris
Farrar, Straus and Giroux, 2022. 336 pp.



To counter the book's unsparing immersion in the atrocities of war, Fitzharris presents vivid depictions of humanity that act as a profound salve. A nurse prepares a glass of warm milk for a hospitalized patient. A woman looks past a soldier's disfigurement and falls in love. Gillies meticulously plans each move in the operating room, committed to the restoration of face and soul, agonizing when things go wrong.

Interwoven through Fitzharris's story is the irony of Gillies's efforts, as encapsulated in an exchange between a bacteriologist and a surgeon in a London hotel lobby overheard by journalist Harold Begbie. "Here we are, you and I, whose business it is to save life, in the midst of men whose business it is to destroy life," the bacteriologist is said to have remarked. There was one bright spot, however, as surgeon Fred Albee observed: "That in the long run, humanity would benefit from the knowledge surgeons had gained in time of war."

Indeed, this would prove to be the case. In addition to pioneering work in reconstructive plastic surgery, battlefield medicine also led to improvements in orthopedic surgery, blood transfusions, and trauma surgery.

After the war, Gillies became the first surgeon to construct a penis for a trans man, thereby pioneering gender affirmation surgery. "The world," his patient wrote after the successful phalloplasty, "began to seem worth living in." Despite negative feedback from some of his peers, Gillies understood that the patient's needs came first.

"The practice of medicine is an art, not a trade; a calling...in which your heart will be exercised equally with your head," William Osler, the father of American medicine, once said. As Fitzharris so elegantly reveals, Harold Gillies embodied this sentiment. ■

The reviewer is a freelance writer and former editor at large at *National Geographic* based in Washington, DC, USA.
Email: cnewmanwriter@gmail.com

10.1126/science.abo7797



Aquatic species risk becoming entangled in fishing nets that have been lost or abandoned.

PHOTO: MARCUS DAVIS ANDRADE BRAGA, MAR DO CEARÁ

LETTERS

Edited by **Jennifer Sills**

New treaty must address ghost fishing gear

In his News story "World's nations start to hammer out first global treaty on plastic pollution" (23 February, https://scim.ag/unplastictreaty), E. Stokstad discusses the issues that may be addressed by a new plastic treaty (*1*), including pollution resulting from fishing activities. Because fishing gear is often made from long-lasting synthetic polymers, such as nylon (*2*), lost and abandoned gear is a long-term problem. This type of pollution, known as ghost gear, is a serious and pervasive threat to the integrity of ecosystems (*2*). The first plastic treaty must address ghost gear in marine (*3*) and freshwater environments.

Ghost gear affects aquatic ecosystems on every continent. Abandoned or lost nets, for example, trap and often kill large fish (e.g., elasmobranchs), crustaceans (decapods), turtles, mammals (including cetaceans), and other organisms (*4*–*7*). Although reports are more frequent from marine ecosystems, damage has occurred in inland water ecosystems as well (*2*, *7*). Other animals, such as birds, are attracted to potential prey trapped in the ghost gear and can become entangled themselves (*5*, *8*), generating a negative cascade effect (*5*). As Stokstad notes, the problem is exacerbated by the lack of reliable data on the frequency and degree of impact of ghost gear in aquatic ecosystems around the world.

Given the increasing demand for resources to feed the world's growing population, fishing will intensify in coming years (*3*, *9*), and the amount of ghost gear in aquatic ecosystems will almost certainly increase as a result. To address this problem, the plastic treaty should aim to reduce the risk fishing gear poses to the environment. Possible strategies include replacing synthetic fishing gear with biodegradable alternatives, which are already available (*10*); limiting the sales of nylon nets; providing educational opportunities; and removing lost and abandoned fishing gear from ecosystems (*2*). In addition to drafting the plastic treaty, all countries must take urgent and comprehensive action to combat the harm caused by fishing activities.

Henrique Vitorino[1], Roberto Ferrazi[1], Guilherme Correia-Silva[1], Felipe Tinti[1], Adrian C. Belizário[1], Fernando A. Amaral[1], Felipe P. Ottoni[2], Carolina V. Silva[1], Tommaso Giarrizzo[3,4], Marlene S. Arcifa[5], Valter M. Azevedo-Santos[1,4,6]*

[1]Faculdade Eduvale de Avaré, Avaré, SP, Brazil. [2]Laboratório de Sistemática e Ecologia de Organismos Aquáticos, Centro de Ciências Agrárias e Ambientais, Universidade Federal do Maranhão, Chapadinha, MA, Brazil. [3]Instituto de Ciências do Mar (LABOMAR), Universidade Federal do Ceará, Fortaleza, Ceará 60165-081, Brazil. [4]Núcleo de Ecologia Aquática e Pesca da Amazônia and Laboratório de Biologia Pesqueira e Manejo dos Recursos Aquáticos, Grupo de Ecologia Aquática, Universidade Federal do Pará, Belém, Pará, Brazil. [5]Departamento de Biologia, Universidade de São Paulo, Ribeirão Preto, SP, Brazil. [6]Programa de Pós-Graduação em Biodiversidade, Ecologia e Conservação, Universidade Federal do Tocantins, Porto Nacional, Tocantins, Brazil.
*Corresponding author. Email: valter.ecologia@gmail.com

REFERENCES AND NOTES

1. UN Environment Assembly of the UN Environment Programme, "End plastic pollution: Towards an international legally binding instrument" (2022); https://wedocs.unep.org/bitstream/handle/20.500.11822/38522/k2200647_-_unep-ea-5-l-23-rev-1_-_advance.pdf?sequence=1&isAllowed=y.
2. V. M. Azevedo-Santos, R. M. Hughes, F. M. Pelicice, *An. Acad. Bras. Ciênc.* **94**, e20201189 (2022).
3. "Stop ghost gear: The most deadly form of marine plastic debris" (World Wildlife Fund, 2020); www.worldwildlife.org/publications/stop-ghost-gear-the-most-deadly-form-of-marine-plastic-debris.
4. J. Adelir-Alves *et al.*, *Braz. J. Oceanogr.* **64**, 427 (2016).
5. V. M. Azevedo-Santos *et al.*, *Mar. Pollut. Bull.* **172**, 112821 (2021).
6. A. J. B. Santos *et al.*, *Herpetol. Rev.* **43**, 245 (2012).
7. V. Iriarte, M. Marmontel, *Aquat. Mamm.* **39**, 116 (2013).
8. T. P. Good *et al.*, *Mar. Ornithol.* **37**, 67 (2009).
9. Food and Agriculture Organization of the United Nations (FAO), "The state of world fisheries and aquaculture" (FAO, Rome, 2014).
10. S. Kim *et al.*, *Anim. Conserv.* **19**, 309 (2016).

10.1126/science.adc9254

Explanations for nitrogen decline

In their Review "Evidence, causes, and consequences of declining nitrogen availability in terrestrial ecosystems" (15 April, eabh3767), R. E. Mason *et al.* argue that nitrogen has decreased in availability worldwide over the past century and that the decline is best explained by human-driven elevated temperatures and CO_2. This conclusion conflicts with previous studies showing strong increases in nitrogen availability compared to preindustrial levels (*1*, *2*). Mason *et al.* present two main types of observational trends as evidence that nitrogen has declined: a decline in Europe and the United States since 1990 in various nitrogen availability indices, and a worldwide decline of nitrogen isotope ratios ($\delta^{15}N$) in plant leaves, tree rings, and lake sediments since 1920. We disagree that rising temperatures and CO_2 levels are the best explanation for these trends.

The decline in nitrogen since 1990 can be easily explained by reduced nitrogen emissions from fossil fuels and agriculture since 1990 in Europe and the United States (*3*). However, because nitrogen emissions remain far above preindustrial levels, high levels of nitrogen inputs in ecosystems continue to cause nitrogen eutrophication and biodiversity loss (*4*). The second trend can be explained by the human-driven shift since 1920 toward a much larger role of gaseous sources of reactive nitrogen in the global nitrogen cycle relative to direct uptake from soils and recycled residues (*1*, *4*). Increasing numbers of livestock, the urine and feces of which contain nitrogen that forms ammonia (NH_3), have led to increased release of this reactive

nitrogen-containing gas into the atmosphere (a process known as volatilization). Artificial nitrogenous fertilizers, which are widely produced from nonreactive nitrogen gas (N_2), have also increased volatilization of nitrogen as ammonia (*5*). Compared with nitrogen released through organic matter decomposition in soils, these gaseous origins of reactive nitrogen are typically more depleted in the stable isotope ^{15}N (*1*, *6*, *7*).

The marked ^{15}N depletion in plants in natural ecosystems over the past century likely reflects these much-increased anthropogenic nitrogen emissions and gases (*6*, *8*, *9*) rather than lower nitrogen availability as Mason *et al.* suggest. Therefore, we caution against Mason *et al.*'s recommendation to fertilize seminatural ecosystems with nitrogen to improve carbon sequestration. To prevent the negative effects of excess nitrogen (such as biodiversity loss), implementing this intervention should wait until more compelling evidence is available.

Han Olff[1]*, Rien Aerts[2], Roland Bobbink[3], J. Hans C. Cornelissen[2], Jan Willem Erisman[4], James N. Galloway[5], Carly J. Stevens[6], Mark A. Sutton[7], Franciska T. de Vries[8], G. W. Wieger Wamelink[9], David A. Wardle[10]

[1]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, Netherlands. [2]Systems Ecology Group, Free University Amsterdam, Amsterdam, Netherlands. [3]B-WARE Research Centre, Radboud University, Nijmegen, Netherlands. [4]Institute of Environmental Sciences, Leiden University, Leiden, Netherlands. [5]Department of Environmental Sciences, University of Virginia, Charlottesville, VA 22903, USA. [6]Lancaster Environment Center, Lancaster University, Landcaster, UK. [7]UK Center for Ecology and Hydrology, Edinburgh, Scotland, UK. [8]Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, Netherlands. [9]Wageningen Environmental Research, Wageningen, Netherlands. [10]Asian School for the Environment, Nanyang Technological University, Singapore.
*Corresponding author. Email: h.olff@rug.nl

REFERENCES AND NOTES

1. D. Fowler *et al.*, *Phil. Trans. R. Soc. B.* **368**, 20130164 (2013).
2. J. N. Galloway, A. Bleeker, J. W. Erisman, *Annu. Rev. Environ. Resour.* **46**, 255 (2021).
3. D. Ackerman, D. B. Millet, X. Chen, *Glob. Biogeochem. Cyc.* **33**, 100 (2019).
4. R. Bobbink *et al.*, *Ecological Applications*. **20**, 30 (2010).
5. B. Pan, S. K. Lam, A. Mosier, D. Chen, *Agricult. Ecosys. Environ.* **232**, 283 (2016).
6. P. D. Erskine *et al.*, *Oecologia* **117**, 187 (1998).
7. N. Bhattarai *et al.*, *Front. Environ. Sci. Eng.* **15**, 126 (2021).
8. G. R. Stewart, M. P. Aidar, C. A. Joly, S. Schmidt, *Oecologia* **131**, 468 (2002).
9. D. M. Vallano, J. P. Sparks, *Oecologia* **172**, 47 (2013).

10.1126/science.abq7575

Response

Olff *et al.* select only a subset of the evidence for declining nitrogen availability and assign unlikely mechanisms to reach the conclusion that nitrogen availability is not declining over large areas of Earth. We disagree that the evidence can be grouped into the categories that Olff *et al.* describe; the complete set of observations is wider in scope and cannot be explained by the mechanisms that the authors propose.

Olff *et al.* claim that declines in nitrogen emissions since 1990 can explain declining nitrogen availability. Our Review acknowledges reduced emissions, and the resulting reduction in atmospheric deposition of nitrogen onto ecosystems, as a likely contributing factor. However, we also present long-term records of declining nitrogen availability, including declining nitrogen concentrations in plant leaves since around 1930 (*1*, *2*) and in plant pollen since the early 1900s (*3*), as well as declines in a broad suite of soil nitrogen availability indicators and stream water NO_3^- at Hubbard Brook in New Hampshire, United States, that date back to the 1960s and 1970s (*4*, *5*). These observations predate reductions in nitrogen deposition. Moreover, as we explain in the Review, declines in nitrogen availability indicators have occurred in places that have never experienced substantially elevated nitrogen deposition (*1*) and alongside declines in concentrations of other elements in plants (*6*–*8*).

Olff *et al.* then propose that large-scale declines in natural abundance nitrogen isotope ratio ($\delta^{15}N$) values in sediment and plants can be explained by a change over time in the isotopic signature of anthropogenic nitrogen emissions toward isotopically lighter, reduced forms of nitrogen. However, the evidence they cite of possible effects of this shift on plant $\delta^{15}N$ refers only to a handful of case studies in atypical environments (*9*–*11*). The isotopic ratio of deposited nitrogen is elevated by processes in soil that discriminate against ^{15}N; the effects of such processes increase with increasing nitrogen supply (*2*, *12*). Models show that the isotopic signature of deposited nitrogen would have to be implausibly low to cause plant $\delta^{15}N$ to decline at the observed rate (*2*).

There is little doubt that massive and poorly managed anthropogenic nitrogen inputs have led to eutrophication and biodiversity loss in many locations. However, rising atmospheric CO_2, warming, and several other global changes are concurrently driving a reduction in nitrogen availability (i.e., nitrogen supply relative to nitrogen demand). The well-documented increases in anthropogenic nitrogen supply noted by Olff *et al.* have not affected global ecosystems uniformly and are unlikely to be the overriding driver of changes in nitrogen availability across all terrestrial ecosystems.

As we state in our Review, the fundamental response to declining nitrogen availability must be to reduce CO_2 emissions. We point out that, although fertilization may be one option for increasing nitrogen availability to plants, microbes, and herbivores, numerous factors must be taken into account when designing interventions that can achieve well-defined goals without unacceptable negative consequences. Further work is necessary to more fully demonstrate the extent of declines in nitrogen availability, to clarify the underlying mechanisms, and to delineate appropriate responses. But before this can happen, the scientific evidence for declining nitrogen availability must be acknowledged.

Rachel E. Mason[1]*, Joseph M. Craine[2], Nina K. Lany[3], Mathieu Jonard[4], Scott V. Ollinger[5], Peter M. Groffman[6,7], Robinson W. Fulweiler[8,9], Jay Angerer[10], Quentin D. Read[11], Peter B. Reich[12,13,14], Pamela H. Templer[9], Andrew J. Elmore[15,16]

[1]Center for Global Discovery and Conservation Science, Arizona State University, Tempe, AZ 85287, USA. [2]Jonah Ventures, Boulder, CO 80301, USA. [3]US Department of Agriculture Forest Service Northern Research Station, Durham, NH 03824, USA. [4]Earth and Life Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium. [5]Earth Systems Research Center, University of New Hampshire, Durham, NH 03824, USA. [6]City University of New York Advanced Science Research Center at the Graduate Center, New York, NY 10031, USA. [7]Cary Institute of Ecosystem Studies, Millbrook, NY 12545, USA. [8]Department of Earth and Environment, Boston University, Boston, MA 02215, USA. [9]Department of Biology, Boston University, Boston, MA 02215, USA. [10]US Department of Agriculture, Agricultural Research Service, Fort Keogh Livestock and Range Research Laboratory, Miles City, MT 59301, USA. [11]US Department of Agriculture Agricultural Research Service, Southeast Area, Raleigh, NC 27695, USA. [12]Department of Forest Resources, University of Minnesota, St. Paul, MN 55108, USA. [13]Institute for Global Change Biology and School for Environment and Sustainability, University of Michigan, Ann Arbor, MI 48109, USA. [14]Hawkesbury Institute for the Environment, Western Sydney University, Penrith, NSW, Australia. [15]Appalachian Laboratory, University of Maryland Center for Environmental Science, Frostburg, MD 21532, USA. [16]National Socio-Environmental Synthesis Center, Annapolis, MD 21401, USA.
*Corresponding author.
Email: rachel.e.mason1@gmail.com

REFERENCES AND NOTES

1. E. N. J. Brookshire, P. C. Stoy, B. Currey, B. Finney, *Glob. Chang. Biol.* **26**, 5404 (2020).
2. K. K. McLauchlan, C. J. Ferguson, I. E. Wilson, T. W. Ocheltree, J. M. Craine, *New Phytol.* **187**, 1135 (2010).
3. L. H. Ziska *et al.*, *Proc. R. Soc. B Biol. Sci.* **283**, 20160414 (2016).
4. J. Durán *et al.*, *Ecosphere* **7**, 1 (2016).
5. P. M. Groffman *et al.*, *Biogeochemistry* **141**, 523 (2018).
6. J. Penuelas *et al.*, *Commun. Biol.* **3**, 1 (2020).
7. M. Jonard *et al.*, *Glob. Chang. Biol.* **21**, 418 (2015).
8. E. A. R. Welti, K. A. Roeder, K. M. De Beurs, A. Joern, M. Kaspari, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 7271 (2020).
9. D. M. Vallano, J. P. Sparks, *Oecologia* **172**, 47 (2013).
10. P. D. Erskine *et al.*, *Oecologia* **117**, 187 (1998).
11. G. R. Stewart, M. P. M. Aidar, C. A. Joly, S. Schmidt, *Oecologia* **131**, 468 (2002).
12. T. E. Dawson, S. Mambelli, A. H. Plamboeck, P. H. Templer, K. P. Tu, *Annu. Rev. Ecol. Syst.* **33**, 507 (2002).

10.1126/science.abq8690

RESEARCH

IN *SCIENCE* JOURNALS

Edited by
Michael Funk



PLANT SCIENCE

Host plant nurtures fungus

Some fungi depend on their living hosts for sustenance. The corn smut fungus *Ustilago maydis* can grow independently but depends on the host maize plant to reproduce. Kretschmer *et al*. analyzed which host nutrients are required to support this obligate biotroph's lifestyle. The fungus responds to a combination of nutrients, including organic acids such as malate, which maize uses as a substrate for C4 photosynthesis. Identification of dicarboxylate transporters showed that the ability of the fungus to draw these organic acid out of the host plant contributes to the pathogen's virulence. With such nutrition ensured, the fungus can then move through its life cycle. —PJH *Science*, abo2401, this issue p. 1187

Corn smut fungus senses nutrients in its maize host that trigger its reproductive cycle.

PHOTO: ARTMARI/SHUTTERSTOCK

COMMUNITY ECOLOGY

More predation in warmer seas

Species richness of many taxa is higher near the equator, and ecologists have long hypothesized that this pattern is linked to stronger interactions between species (e.g., competition and predation) in the tropics. However, empirical evidence showing that the strength of species interactions varies with latitude is limited. Ashton *et al*. tested whether predation on benthic marine communities is higher at lower latitudes. Using a standardized experiment at 36 sites along the Pacific and Atlantic coasts of North and South America, the authors found both greater predation intensity (consumption rate) and stronger impacts on benthic communities nearer the equator. These trends were more strongly related to water temperature than to latitude, suggesting that climate warming may influence top-down control of communities. —BEL

Science, abc4916, this issue p. 1215

MICROBIOME

Establishing early diversity

Humans living an urbanized lifestyle in industrialized countries tend to have less diverse microbiota than people living more rural existences. Using fecal 16S ribosomal RNA sequencing, Olm *et al*. found that after the first 6 months of life, the microbiome of infants living in contrasting environments diverged from *Bifidobacteria*-dominated assemblages. Deep metagenomic sequencing revealed that a large proportion of the bacterial species detected in samples from hunter-gatherer infants were new and were undetectable in samples from urbanized children. Gut microbiota diversity appears early in the lives of hunter-gatherer infants and is traceable to maternal transmission, with some influence from the local environment. The main driver for differences among gut microbiota originates in lifestyle rather than geography. It is suspected, but still enigmatic, that such differences in microbiota have functional implications for the health of developing children. —CA

Science, abj2972, this issue p. 1220

ORGANIC CHEMISTRY

Catalysis paired with crystallization

Asymmetric catalysis often distinguishes mirror-image configurations at a single carbon center. However, many complex molecules have three or more chiral centers, and selecting just one of the ballooning number of diastereomers in such cases can be daunting. In this context, de Jesús Cruz *et al*. report that product crystallization during the reaction can supplement the catalyst's intrinsic selectivity. Specifically, the authors used a chiral base to set one stereocenter in a Michael addition of nitroalkanes to ketoamides while dynamically scrambling the configurations on the adjacent carbons. Crystallization then selects a single diastereomer from this interconverting mixture. —JSY

Science, abo5048, this issue p. 1224

AGING

Timing eating and fasting for longevity

Animals fed a limited number of calories, just enough to avoid malnutrition, show extended health

span and life span. However, they are so hungry that they eat those fewer calories in a limited period of time and consequently spend more time fasting than do animals for which access to food is not restricted. Acosta-Rodriguez *et al*. therefore designed experiments in mice to control both caloric intake and the timing of their eating to see which factors were the most important (see the Perspective by Deota and Panda). Caloric restriction extended life span as expected, but it worked best when feeding was restricted so that the animals fasted for at least 12 hours and when the period in which the animals ate corresponded to the active phase of their circadian cycle. —LBR

Science, abk0297, this issue p. 1192; see also adc8824, p. 1159

MARINE VIROME

Patterns and process in RNA viruses

Viruses are suspected to be lynchpins in ecosystem function, but so far we can only guess at their significance. DNA viruses are increasingly being recognized as significant components of biogeochemical cycling in the oceans. Dominguez-Huerta *et al*. explored global patterns of marine RNA virus occurrence by extracting virus sequences from Tara Ocean samples. Host prediction analysis identified predominantly protist and fungal hosts plus a few invertebrates. Like double-stranded DNA viruses and their hosts, RNA viruses showed marked depth limitation but little latitudinal change. Auxiliary metabolic genes in the RNA virome indicated that several eukaryote plankton processes are affected by viruses. A group of 11 RNA viruses that significantly influence ocean carbon flux were identified. —CA

Science, abn6358, this issue p. 1202

PARKINSON'S DISEASE

Big step forward for Parkinson's disease

Inhibition of the kinase LRRK2 has emerged as a promising disease-modifying therapeutic target for Parkinson's disease. Jennings *et al*. report evidence that DNL201, a first-in-class central nervous system–penetrant LRRK2 kinase inhibitor, reduces LRRK2 activity and restores lysosomal function in cellular and animal models. In healthy volunteers and patients with Parkinson's disease, DNL201 inhibited LRRK2 kinase activity and demonstrated an impact on lysosomal function at doses that were safe and generally well tolerated. These findings provide support for advancing the investigation of LRRK2 inhibitors to late-stage clinical studies in patients with Parkinson's disease (see the Focus by Lewis). —OMS

Sci. Transl. Med. **14**, eabj2658 (2022); see also abq7374 (2022).



Surveys from the Tara Ocean's research vessel, pictured here, provide insight into global RNA virus abundance and distribution.

IN OTHER JOURNALS

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



REPRODUCTIVE HEALTH

Distant access to abortion

The US Supreme Court is considering overturning *Roe v Wade*, a 1973 ruling that legalized abortion across the country. If overturned, the decision to maintain this essential health service will be left to individual states. One consequence for patients living in southern conservative states will be long-distance travel to abortion facilities in more liberal northern and western states. Pleasants *et al*. used the Google Ads Abortion Access Study to examine pregnancy outcomes for individuals considering an abortion. At 4-week follow-up, women who lived farther from an abortion facility (50+ miles) had significantly higher odds of still being pregnant because of the prohibitive travel costs of traveling to a clinic. A lack of local abortion access could be far-reaching for patients with underlying medical conditions, and others will suffer from the economic and mental health burdens of continuing unwanted pregnancies. —EEU

JAMA Netw. Open **5**, e2212065 (2022).

METABOLISM

Fast heat

Mammals, including humans, have two types of body fat named after their color: white fat and brown fat. Most fat in humans is white fat, which stores energy. By contrast, brown fat breaks down sugar and lipids to generate heat when we are exposed to cold temperatures. The thermogenic process triggers rapidly and happens partly through the activation of a gene called uncoupling protein 1 (*Ucp1*), which uncouples protons moving down a mitochondrial gradient from ATP synthesis, thus allowing the energy to be dissipated as heat. Wang *et al*. have evidence to explain why brown fat can produce heat so quickly. A protein called DDB1 occupies the promoters of many thermogenic genes, including *Ucp1*, and releases a brake on

PHOTOS: (LEFT TO RIGHT) @VINCENT HILAIRE; KATYA TSVETKO

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

QUANTUM COMPUTING

Learning from quantum experiments

There is considerable interest in extending the recent success of quantum computers in outperforming their conventional classical counterparts (quantum advantage) from some model mathematical problems to more meaningful tasks. Huang *et al*. show how manipulating multiple quantum states can provide an exponential advantage over classical processing of measurements of single-quantum states for certain learning tasks. These include predicting properties of physical systems, performing quantum principal component analysis on noisy states, and learning approximate models of physical dynamics (see the Perspective by Dunjko). In their proof-of-principle experiments using up to 40 qubits on a Google Sycamore quantum processor, the authors achieved almost four orders of magnitude of reduction in the required number of experiments over the best-known classical lower bounds. —YS

Science, abn7293, this issue p. 1182; see also abp9885, p. 1154

QUANTUM SIMULATION

Solving hard graph problems

Realizing quantum speedup for solving practical, computationally hard problems is the central challenge in quantum information science. Ebadi *et al*. used Rydberg atom arrays composed of up to 289 coupled qubits in two spatial dimensions to investigate quantum optimization algorithms for solving the maximum independent set, a paradigmatic nondeterministic polynomial time–hard combinatorial optimization problem (see the Perspective by Schleier-Smith). A hardware-efficient encoding protocol associated with Rydberg blockade was used to realize a closed-loop optimization method to test several variational algorithms and subsequently apply them to systematically explore a class of nonplanar graphs with programmable connectivity. The results demonstrate the potential of quantum machines as a tool for the discovery of new promising algorithm classes. —ISO

Science, abo6587, this issue p. 1209; see also abq3754, p. 1155

PHOSPHORUS CHEMISTRY

Chiral-at-P products via H-bond catalysis

Hydrogen (H)–bonding catalysis has recently proven useful for activating carbon-chlorine bonds to form just one of two possible mirror-image products. Forbes and Jacobsen now extend this approach to desymmetrization of phosphorus(V) [P(V)] dichloride compounds (see the Perspective by Verdaguer). Using chiral urea catalysts, the authors could displace just one of two chlorides with an amine, thereby producing a versatile P(V) intermediate. Subsequent selective displacement of the remaining chloride and/or amine offers access to a wide range of chiral-at-P compounds, a class of increasing pharmaceutical interest. —JSY

Science, abp8488, this issue p. 1230; see also abq5073, p. 1157

PROTEIN BIOPHYSICS

Stabilizing receptors with cholesterol

Membrane-embedded G protein–coupled receptors such as the β_2-adrenergic receptor (β_2AR) mingle with cholesterol, which modulates their assembly and stability. Serdiuk *et al*. examined the effects of a cholesterol analog on the mechanical and energetic properties of the β_2AR at different temperatures. The authors showed that whereas the presence of the cholesterol analog in liposomes had minimal effects at 25° and 42°C, it stabilized subsets of conformations of the β_2AR that were important for basal activity at 37°C. —JFF

Sci. Signal. **15**, eabi7031 (2022).

CANCER IMMUNOLOGY

ILC2s help tumors

Group 2 innate lymphoid cells (ILC2s) protect against parasitic infections and are implicated in allergies, yet their role in antitumor immunity remains unclear. Jou *et al.* used various genetic mouse models to study the roles of ILC2s in colorectal cancer (CRC). They found that ILC2s were linked to an immunosuppressive tumor microenvironment, where deletion of these cells led to less CRC tumor burden. ILC2s specifically responded to high interleukin-25 (IL-25) expression in CRC tumors, and in turn inducing immunosuppressive myeloid-derived suppressor cells. Therapeutically blocking the IL-25 receptor on ILC2s lowered tumor burden and led to more favorable antitumor immune responses. —DAE

Sci. Immunol. **7**, eabn0175 (2022).

span and life span. However, they are so hungry that they eat those fewer calories in a limited period of time and consequently spend more time fasting than do animals for which access to food is not restricted. Acosta-Rodriguez *et al*. therefore designed experiments in mice to control both caloric intake and the timing of their eating to see which factors were the most important (see the Perspective by Deota and Panda). Caloric restriction extended life span as expected, but it worked best when feeding was restricted so that the animals fasted for at least 12 hours and when the period in which the animals ate corresponded to the active phase of their circadian cycle. —LBR

Science, abk0297, this issue p. 1192; see also adc8824, p. 1159

MARINE VIROME

Patterns and process in RNA viruses

Viruses are suspected to be lynchpins in ecosystem function, but so far we can only guess at their significance. DNA viruses are increasingly being recognized as significant components of biogeochemical cycling in the oceans. Dominguez-Huerta *et al*. explored global patterns of marine RNA virus occurrence by extracting virus sequences from Tara Ocean samples. Host prediction analysis identified predominantly protist and fungal hosts plus a few invertebrates. Like double-stranded DNA viruses and their hosts, RNA viruses showed marked depth limitation but little latitudinal change. Auxiliary metabolic genes in the RNA virome indicated that several eukaryote plankton processes are affected by viruses. A group of 11 RNA viruses that significantly influence ocean carbon flux were identified. —CA

Science, abn6358, this issue p. 1202

PARKINSON'S DISEASE

Big step forward for Parkinson's disease

Inhibition of the kinase LRRK2 has emerged as a promising disease-modifying therapeutic target for Parkinson's disease. Jennings *et al*. report evidence that DNL201, a first-in-class central nervous system–penetrant LRRK2 kinase inhibitor, reduces LRRK2 activity and restores lysosomal function in cellular and animal models. In healthy volunteers and patients with Parkinson's disease, DNL201 inhibited LRRK2 kinase activity and demonstrated an impact on lysosomal function at doses that were safe and generally well tolerated. These findings provide support for advancing the investigation of LRRK2 inhibitors to late-stage clinical studies in patients with Parkinson's disease (see the Focus by Lewis). —OMS

Sci. Transl. Med. **14**, eabj2658 (2022); see also abq7374 (2022).



Surveys from the Tara Ocean's research vessel, pictured here, provide insight into global RNA virus abundance and distribution.

IN OTHER JOURNALS

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



REPRODUCTIVE HEALTH

Distant access to abortion

The US Supreme Court is considering overturning *Roe v Wade*, a 1973 ruling that legalized abortion across the country. If overturned, the decision to maintain this essential health service will be left to individual states. One consequence for patients living in southern conservative states will be long-distance travel to abortion facilities in more liberal northern and western states. Pleasants *et al*. used the Google Ads Abortion Access Study to examine pregnancy outcomes for individuals considering an abortion. At 4-week follow-up, women who lived farther from an abortion facility (50+ miles) had significantly higher odds of still being pregnant because of the prohibitive travel costs of traveling to a clinic. A lack of local abortion access could be far-reaching for patients with underlying medical conditions, and others will suffer from the economic and mental health burdens of continuing unwanted pregnancies. —EEU

JAMA Netw. Open **5**, e2212065 (2022).

METABOLISM

Fast heat

Mammals, including humans, have two types of body fat named after their color: white fat and brown fat. Most fat in humans is white fat, which stores energy. By contrast, brown fat breaks down sugar and lipids to generate heat when we are exposed to cold temperatures. The thermogenic process triggers rapidly and happens partly through the activation of a gene called uncoupling protein 1 (*Ucp1*), which uncouples protons moving down a mitochondrial gradient from ATP synthesis, thus allowing the energy to be dissipated as heat. Wang *et al*. have evidence to explain why brown fat can produce heat so quickly. A protein called DDB1 occupies the promoters of many thermogenic genes, including *Ucp1*, and releases a brake on

PHOTOS: (LEFT TO RIGHT) @VINCENT HILAIRE; KATYA TSVETKO



EVOLUTION

Cetacean skulls from land to ocean

Cetaceans have populated the ocean to become charismatic subjects of human legend, but it has taken 50 million years for whales and dolphins to evolve from terrestrial mammals into an array of specialized aquatic species. Their shift to water required enormous physiological and physical adaptations. Specializations such as alterations in the posterior skull to accommodate breathing at the water's surface, baleen for filter feeding, and biosonar for echolocation are hallmarks of cetacean adaptations. Coombs *et al.* analyzed in three dimensions the shapes of skulls from hundreds of living and extinct cetacean species. Their results show that rapid evolution during the Eocene reshaped the front of the skull. Further diversification of skull morphology continued as cetaceans evolved to accommodate new diets and feeding methods, as well as echolocation. —PJH *Curr. Biol.* **32**, 2233 (2022).

Since the Eocene, when their ancestors re-entered the sea, skulls of cetaceans evolved rapidly to adapt to a subaquatic lifestyle.

the transcription of these genes, allowing them to be expressed. Increased understanding of the mechanism is important because brown fat thermogenesis is crucial not only for maintaining core body temperature but also for energy balance. Some studies have explored the potential of brown fat in treating obesity. —DJ

Life Metab. 10.1093/lifemeta/loac003 (2022).

NEUROSCIENCE

Flexible learning in the cerebellum

The cerebellum is known to participate in the acquisition of motor skills and motor learning, but it also processes reward-related information. The interaction between these pathways and their contribution to learning is not understood. Sendhilnathan *et al.* compared the activity of cerebellar Purkinje cells while monkeys were actively learning new visuomotor associations and while they performed an already familiar visuomotor association task. These cells process two different signals when monkeys learn a new arbitrary stimulus–response association. One is a previously described reinforcement learning error signal, and the second, described here, is a signal that describes the state of learning. —PRS

J. Neurosci. **42**, 3847 (2022).

VOLCANOLOGY

Eruption forerunners

Volcanic eruptions often have different sorts of precursors before the main event that range from a change in the number of small earthquakes to the composition of gases from fumaroles. Flóvenz *et al.* show that pre-eruptive uplift cycles occurred in a nearby geothermal field before the March 2021 Fagradalsfjall eruption in Iceland. The changes were likely due to fluid migration from the magma body and occurred up to a year and a half before the eruption. The observations help provide different constraints on parts of the complex process that can lead to a volcanic eruption. —BG

Nat. Geosci. **15**, 397 (2022).

NONLINEAR OPTICS

Relaxing constraints for phase matching

The interactions of light in nonlinear optical materials produce a number of effects, including lasing, frequency conversion, high-harmonic generation, and spontaneous downconversion. Such effects find application in microscopy, optical communication networks, and quantum optics. Underlying these effects is phase matching of the interacting light beams, which requires strict experimental conditions to be met, often leading to cumbersome setups. Gagnon *et al.* show that metamaterials engineered with a low refractive index can relax the constraints for phase matching. Their demonstration of direction-independent four-wave mixing in a nanophotonic structure illustrates how such low-index materials can be used to miniaturize nonlinear optical devices. — ISO

Phys. Rev. Lett. **128**, 203902 (2022).

CHEMISTRY

Molecules work on self-control

Oscillation of a chemical system between two or more states underlies many important processes in biology and may be a useful model for understanding how early cells formed and functioned. Howlett *et al.* designed a model supramolecular system that achieves a rudimentary metabolism: oscillatory self-replication of a metastable micelle powered by a chemical fuel. The micelle contains a disulfide that can cross-react with other components, leading to dispersal and eventual regeneration. The addition of a dye that is incorporated into the transient micelle provides a visual readout of the supramolecular oscillations. —MAF

Nat. Chem. 10.1038/s41557-022-00949-6 (2022).

RESEARCH ARTICLE SUMMARY

NUCLEAR PORE COMPLEX

Structure of the cytoplasmic ring of the *Xenopus laevis* nuclear pore complex

Xuechen Zhu†, Gaoxingyu Huang*†, Chao Zeng†, Xiechao Zhan†, Ke Liang†, Qikui Xu, Yanyu Zhao, Pan Wang, Qifan Wang, Qiang Zhou, Qinghua Tao, Minhao Liu, Jianlin Lei, Chuangye Yan, Yigong Shi*

INTRODUCTION: The nuclear pore complex (NPC) resides on the nuclear envelope (NE) and mediates nucleocytoplasmic cargo transport. As one of the largest cellular machineries, a vertebrate NPC consists of cytoplasmic filaments, a cytoplasmic ring (CR), an inner ring, a nuclear ring, a nuclear basket, and a luminal ring. Each NPC has eight repeating subunits. Structure determination of NPC is a prerequisite for understanding its functional mechanism. In the past two decades, integrative modeling, which combines x-ray structures of individual nucleoporins and subcomplexes with cryo–electron tomography reconstructions, has played a crucial role in advancing our knowledge about the NPC.

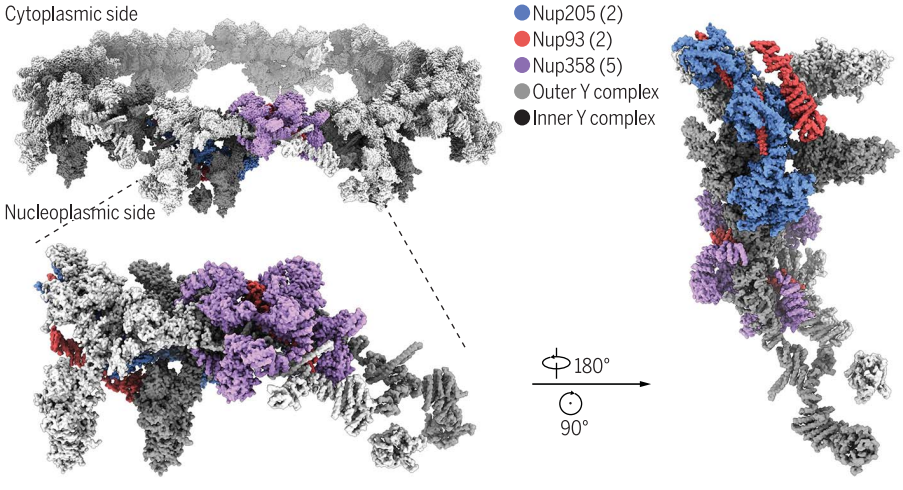
The CR has been a major focus of structural investigation. The CR subunit of human NPC was reconstructed by cryo–electron tomography through subtomogram averaging to an overall resolution of ~20 Å, with local resolution up to ~15 Å. Each CR subunit comprises two Y-shaped multicomponent complexes known as the inner and outer Y complexes. Eight inner and eight outer Y complexes assemble in a head-to-tail fashion to form the proximal and distal rings, respectively, constituting the CR scaffold. To achieve higher resolution of the CR, we used single-particle cryo–electron microscopy (cryo-EM) to image the intact NPC from the NE of *Xenopus laevis* oocytes. Reconstructions of the core region and the Nup358 region of the *X. laevis* CR subunit had been achieved at average resolutions of 5 to 8 Å, allowing identification of secondary structural elements.

RATIONALE: Packing interactions among the components of the CR subunit were poorly defined by all previous EM maps. Additional components of the CR subunit are strongly suggested by the EM maps of 5- to 8-Å resolution but remain to be identified. Addressing these issues requires improved resolution of the cryo-EM reconstruction. Therefore, we may need to enhance sample preparation, optimize image acquisition, and develop an effective data-processing strategy.

RESULTS: To reduce conformational heterogeneity of the sample, we spread the opened NE onto the grids with minimal force and used the chemical cross-linker glutaraldehyde to stabilize the NPC. To alleviate orientation bias of the NPC, we tilted sample grids and imaged the sample with higher electron dose at higher angles. We improved the image-processing protocol. With these efforts, the average resolutions for the core and the Nup358 regions have been improved to 3.7 and 4.7 Å, respectively. The highest local resolution of the core region reaches 3.3 Å. In addition, a cryo-EM structure of the N-terminal α-helical domain of Nup358 has been resolved at 3.0-Å resolution. These EM maps allow the identification of five copies of Nup358, two copies of Nup93, two copies of Nup205, and two copies of Y complexes in each CR subunit. Relying on the EM maps and facilitated by AlphaFold prediction, we have generated a final model for the CR of the *X. laevis* NPC. Our model of the CR subunit includes 19,037 amino acids in 30 nucleoporins.

A previously unknown C-terminal fragment of Nup160 was found to constitute a key part of the vertex, in which the short arm, long arm, and stem of the Y complex meet. The Nup160 C-terminal fragment directly binds the β-propeller proteins Seh1 and Sec13. Two Nup205 molecules, which do not contact each other, bind the inner and outer Y complexes through distinct interfaces. Conformational elasticity of the two Nup205 molecules may underlie their versatility in binding to different nucleoporins in the proximal and distal CR rings. Two Nup93 molecules, each comprising an N-terminal extended helix and an ACE1 domain, bridge the Y complexes and Nup205. Nup93 and Nup205 together play a critical role in mediating the contacts between neighboring CR subunits. Five Nup358 molecules, each in the shape of a shrimp tail and named "the clamp," hold the stems of both Y complexes. The innate conformational elasticity allows each Nup358 clamp to adapt to a distinct local environment for optimal interactions with neighboring nucleoporins. In each CR subunit, the α-helical nucleoporins appear to provide the conformational elasticity; the 12 β-propellers may strengthen the scaffold.

CONCLUSION: Our EM map–based model of the *X. laevis* CR subunit substantially expands the molecular mass over the reported composite models of vertebrate CR subunit. In addition to the Y complexes, five Nup358, two Nup205, and two Nup93 molecules constitute the key components of the CR. The improved EM maps reveal insights into the interfaces among the nucleoporins of the CR. ■



Cryo-EM structure of the double-layered CR of the *X. laevis* NPC. The *X. laevis* CR, containing eight repeating subunits, is modeled on the basis of cryo-EM reconstructions (top left panel). One CR subunit is shown in two different views to highlight nucleoporins of key interest (bottom left and right panels). The inner and outer Y complexes are colored dark and light gray, respectively. Two Nup205, two Nup93, and five Nup358 molecules are colored blue, red, and purple, respectively.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: huanggaoxingyu@westlake.edu.cn (G.H.); syg@westlake.edu.cn (Y.S.)
†These authors contributed equally to this work.
Cite this article as X. Zhu *et al.*, *Science* **376**, eabl8280 (2022). DOI: 10.1126/science.abl8280

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abl8280

RESEARCH ARTICLE

NUCLEAR PORE COMPLEX

Structure of the cytoplasmic ring of the *Xenopus laevis* nuclear pore complex

Xuechen Zhu[1,2,3]†, Gaoxingyu Huang[1,2,3]*†, Chao Zeng[4,5]†, Xiechao Zhan[1,2,3]†, Ke Liang[1,2,3]†, Qikui Xu[1,2,3], Yanyu Zhao[1,2,3], Pan Wang[4,5], Qifan Wang[1,2,3], Qiang Zhou[1,2,3], Qinghua Tao[4], Minhao Liu[4], Jianlin Lei[4], Chuangye Yan[4,5], Yigong Shi[1,2,3,4,5]*

The nuclear pore complex (NPC) mediates nucleocytoplasmic cargo transport. Here, we present a single-particle cryo–electron microscopy reconstruction of the cytoplasmic ring (CR) subunit from the *Xenopus laevis* NPC at 3.7- to 4.7-angstrom resolution. The structure of an amino-terminal domain of Nup358 has been resolved at 3.0 angstroms, facilitating the identification of five Nup358 molecules in each CR subunit. Our final model of the CR subunit included five Nup358, two Nup205, and two Nup93 molecules in addition to the two previously characterized Y complexes. The carboxyl-terminal fragment of Nup160 served as an organizing center at the vertex of each Y complex. Structural analysis revealed how Nup93, Nup205, and Nup358 facilitate and strengthen the assembly of the CR scaffold that is primarily formed by two layers of Y complexes.

The nuclear pore complex (NPC) resides on the nuclear envelope (NE) and mediates nucleocytoplasmic cargo transport (*1*–*3*). As one of the largest cellular machineries (*4*–*6*), a vertebrate NPC has a molecular mass of >100 MDa and consists of multiple cytoplasmic filaments (CF), a cytoplasmic ring (CR), an inner ring (IR), a nuclear ring (NR), a nuclear basket (NB), and a luminal ring (LR) (*5*–*9*) (Fig. 1A). Each NPC has eight repeating subunits known as the spokes (*7*). Our present study concerns the structure of the CR constituents in each spoke, referred to as the CR subunits.

For structural elucidation of the NPC, cryo–electron tomography used to be the mainstream approach. The CR subunit of the human NPC was reconstructed through subtomogram averaging to a highest resolution of ~15 Å (*10*). Each CR subunit is featured with two Y-shaped multicomponent complexes known as the inner and the outer Y complexes. Sixteen Y complexes, eight inner and eight outer, in the CR assemble in a head-to-tail fashion to build the scaffold in the form of the proximal and distal concentric ring, respectively (Fig. 1, A and B) (*11*, *12*). The tripartite Y complex consists of a short arm (Nup85, Nup43, and Seh1), a long arm (Nup160 and Nup37), and a stem (Nup96, Sec13, Nup107, and Nup133) (*11*, *13*) (Fig. 1, A and D).

[1]Westlake Laboratory of Life Sciences and Biomedicine, Westlake University, 310024 Hangzhou, China. [2]Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, 310024 Hangzhou, China. [3]Institute of Biology, Westlake Institute for Advanced Study, 310024 Hangzhou, China. [4]Beijing Advanced Innovation Center for Structural Biology and Frontier Research Center for Biological Structure, Tsinghua University, 100084 Beijing, China. [5]Tsinghua University-Peking University Joint Center for Life Sciences, School of Life Sciences, Tsinghua University, 100084 Beijing, China.
*Corresponding author. Email: huanggaoxingyu@westlake.edu.cn (G.H.); syg@westlake.edu.cn (Y.S.)
†These authors contributed equally to this work.

To achieve higher resolution of the CR assembly, we used single-particle cryo–electron microscopy (cryo-EM) reconstruction to examine the CR subunit from the *Xenopus laevis* NPC. During data processing, two relatively independent and rigid structural entities stood out, designated as the core region and the Nup358 region (*14*). The core region contains the long arm from the inner Y complex and the short arms from both Y complexes, and the Nup358 region covers the stems of the two Y complexes (Fig. 1B). An overall resolution of 5 to 8 Å was achieved after masked refinement for these two regions, enabling assignment of some components, such as two Nup205 and two Nup358 molecules (*14*). However, reliable docking and modeling requires higher resolution.

In this study, we achieved an average resolution of 3.7 Å for the core region and 4.7 Å for the Nup358 region. We also solved a cryo-EM structure of the N-terminal α-helical domain of Nup358 at 3.0-Å resolution. Relying on the improved resolutions and facilitated by AlphaFold prediction, we were able to generate a structural model for the entire CR of the *X. laevis* NPC. In addition to the previously characterized Y complexes, five Nup358, two Nup205, and two Nup93 molecules were identified to be constituents of the CR subunit. Our current structure expands the molecular mass by 80% compared with the previously reported composite models of the human CR (*10*, *12*).

Overall structure of a CR subunit

To reduce conformational heterogeneity, we spread the opened NE onto the grids with minimal force and used the chemical cross-linker glutaraldehyde to stabilize the NPC. Please refer to the materials and methods, figs. S1 to S4, and table S1 for details of cryo-EM sample preparation, image acquisition, data processing, and structural modeling and refinement. We previously defined the core and the Nup358 regions in each CR unit based on their spatial arrangement but were unable to reliably assign nucleoporins (Fig. 1B) (*14*). The average resolutions for these two regions now reach 3.7 and 4.7 Å, with the highest local resolution of the core region up to 3.3 Å (Fig. 1C, Movie 1, and fig. S5).

Structural modeling for the best-resolved components within the CR subunit, including the C-terminal fragment of Nup160 (Nup160-CTF), Nup93 α5, Nup205, and other previously recognized structural components of the CR, was aided by x-ray structures (*12*) or AlphaFold predictions (*15*). Assignment of the two Nup93 α5 helices was facilitated by several bulky side chains that are readily discernible in the EM map. In addition, the model of this long helix predicted by AlphaFold matches the EM map exceptionally well (fig. S5). Assignment of the residues in each *X. laevis* nucleoporin was assisted by sequence alignment with its functional orthologs. The EM density in the Nup358 region and the periphery of the core region, including all five Nup358 clamps and two Ancestral Coatomer Element 1 (ACE1) domains that belong to Nup93 (*16*), is less well resolved and lacks side chain features. Modeling in these areas was aided by the separately determined atomic structure of Nup358-NTD2 (residues 222 to 738) (figs. S6 to S9), previous analyses of vertebrate NPC (*14*), and AlphaFold predictions (*15*).

Altogether, our final model of the CR subunit consists of 19,037 amino acids in 30 nucleoporins (Fig. 1, A and D, Movie 1, and tables S2 and S3). Except for Nup358, Nup88, Nup155, and Nup98, all protein components in the current model have two distinct copies in each CR subunit (*14*). For description clarity, the copy closer to and away from the central channel will be denoted with -I (inner) and -O (outer), respectively. In the following sections, we will focus on previously uncharacterized structural features of the CR for illustration.

Nup160-CTF at the vertex of the Y complex

The Y complexes have been the most extensively characterized, with most of their constituents known. On the basis of the improved EM map of the core region, the previously unknown Nup160-CTF, comprising helices α38 to α47, was found to sit at the center of the vertex, where the short arm, long arm, and stem of the Y complex meet (*11*, *17*) (Fig. 2, A and B). This observation helps to define the organization of the vertex in vertebrates.

Seh1-I from the short arm and Sec13-I from the stem are connected to the CTF of Nup160-I (Fig. 2, A and C). For the inner Y complex, the 4CD loop (the loop between strands C and D of blade 4) of Seh1 contacts helix α40 of Nup160-CTF, whereas the β-3D strand (the D strand of blade 3) of Seh1 interacts with helix α42 of

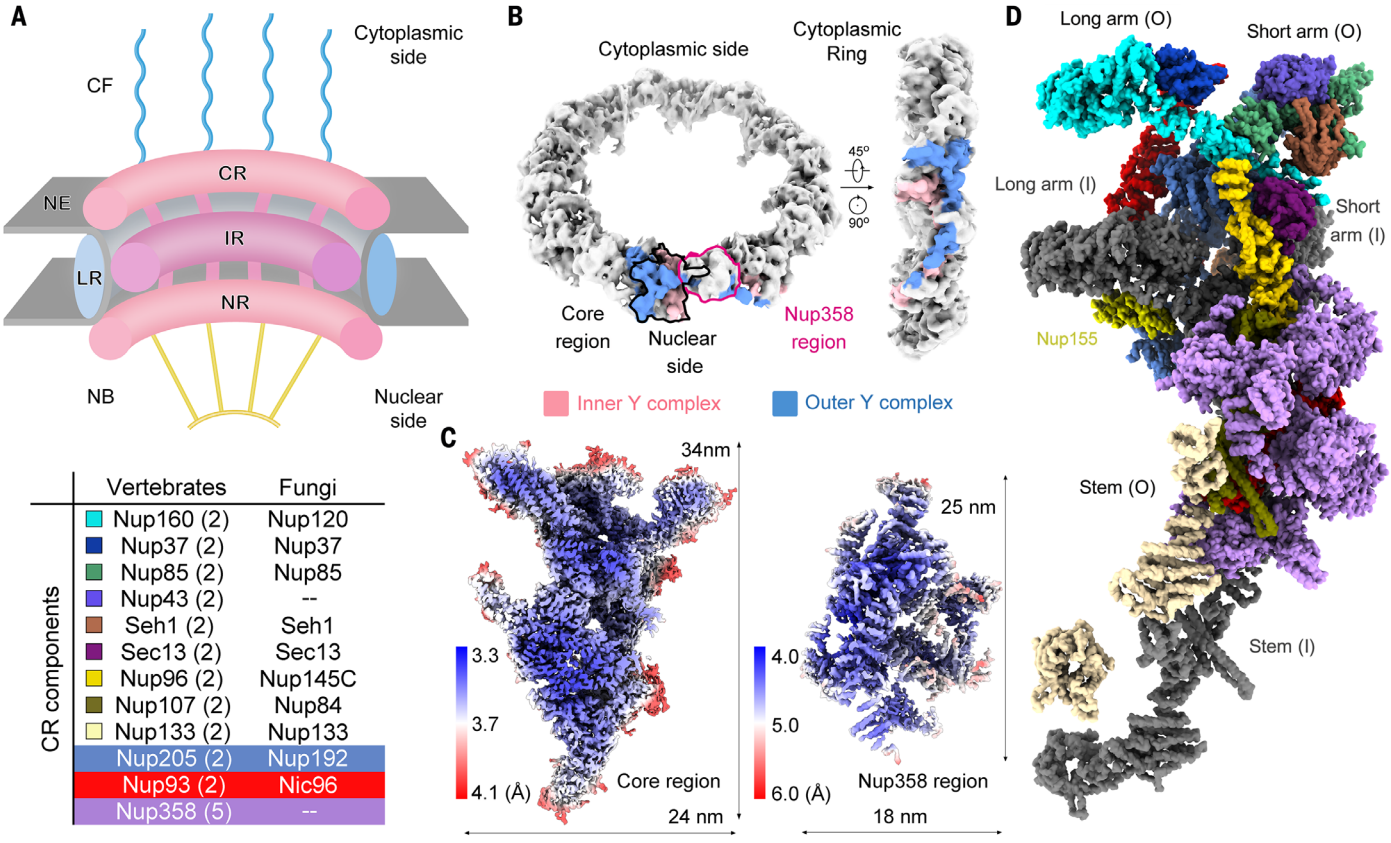


Fig. 1. Single-particle cryo-EM analysis of the CR subunit of the *X. laevis* NPC. (**A**) Schematic overview of the architecture of a vertebrate NPC. Shown is a cut-open side view of a cartoon depicting an NPC embedded in the NE. Each NPC, along its cytonuclear axis, comprises the CF, CR, IR, NR and NB. A LR localized inside the lumen of the NE encompasses the IR. The CR components from vertebrates and fungi are listed in the table below. The copy number of each CR component in a vertebrate NPC is indicated in brackets. (**B**) EM map of the CR at 22.2-Å resolution. The core region and the Nup358 region in one CR subunit, which are separately calculated as single particles, are indicated by black and magenta contours, respectively, in the left panel. The densities corresponding to the inner and outer Y complexes in one CR subunit are colored pink and blue, respectively. (**C**) Heatmaps for the local resolutions of the core region and the Nup358 region that were separately reconstructed to average resolutions of 3.7 and 4.7 Å, respectively. The local resolutions were calculated in RELION3.0 (*59*) and presented in ChimeraX (*58*). (**D**) Overall structure of a CR subunit from the *X. laevis* NPC. The structure is presented in approximately the same view as the two Y complexes in the right panel of (B). Protein components of the outer Y complex and Nup205, Nup93, and Nup358 molecules are colored the same as those annotated in (A). The inner Y complex is colored dark gray. Nup155, which connects the CR to the IR, is labeled but not included in the table in (A).

Nup160-CTF (Fig. 2C). The N-terminal α1 helices from Nup96 and Sec13 pair up to wedge into the crevice between helices α46 and α47 of inner Nup160 (Fig. 2C). These interactions may stabilize the Y complex.

A conserved surface loop between α36 and α37 of Nup160-I, referred to as the 96-loop, closely associates with the ridge of the C-terminal helices from Nup96-I (Fig. 2D). Similar to that in fungi (*18*, *19*), two C-terminal helices of Nup85-I contact the lateral side of the middle portion of the α-helical domain of Nup160-I (Fig. 2D). The interactions described here for inner Y complex are also recapitulated in the outer Y complex.

Two Nup205 molecules bind the inner and outer Y complexes through distinct interfaces

Our structure identifies two Nup205 molecules, defined as Nup205-I and Nup205-O, in each CR subunit (Fig. 3A and fig. S5). The two Nup205, which do not contact each other, associate with the inner and outer Y complexes through different interfaces (Fig. 3A).

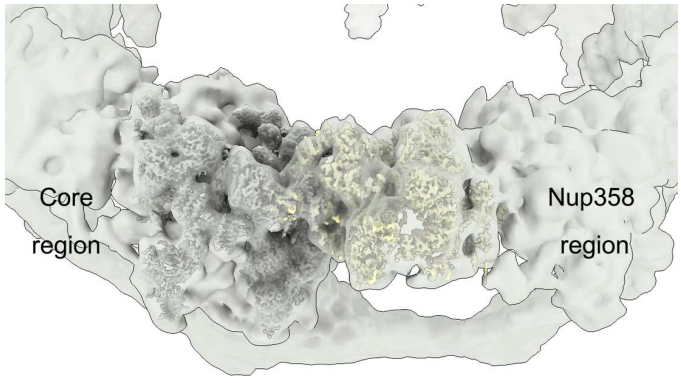
Nup205-O directly contacts the short arm and the vertex of outer Y complex (Fig. 3B). On one side, the tower helix (*12*) and six additional helices associate with the ridge of Nup85-O in the short arm. On the other side, five helices contact Nup160-CTF-O at the vertex (Fig. 3B). Nup205-O also associates with both arms of the inner Y complex through interactions with the bottom faces of the β-propellers in Nup43-I and Nup37-I (Fig. 3C). Through these interactions, Nup205-O connects the two Y complexes within the same CR subunit.

The interactions of Nup205-I with the middle helices of Nup85-I and Nup160-CTF-I are generally conserved as their outer counterparts (Fig. 3, D to F). However, there is no contact between the C-terminal helices of Nup205-I and Nup85-I, which are separated by the N-terminal propeller of Nup88 and Nup98/X from the Nup214 complex (Fig. 3E) (*14*).

These structural observations uncover previously unrecognized interfaces in the organization of the CR scaffold and identify Nup205 as a new core component of the CR. Distinct conformations of the two Nup205 molecules demonstrate the structural elasticity that underlies their versatility for binding to different nucleoporins in the proximal and distal CR rings and the cytoplasmic filaments.

Nup93 bridges the Y complexes and Nup205

Nup93 consists of an N-terminal extended α-helix (α5) and an ACE1 domain that are connected by flexible linkers (fig. S10) (*20*). Our EM maps unveil two Nup93-ACE1. The predicted structure of helix α5 of Nup93 fits exceptionally well into our EM density that is



Movie 1. Single-particle cryo-EM analysis of the CR of the *X. laevis* NPC. The 70-s movie shows the three-dimensional EM reconstructions of the CR (1 to 52 s) and the surface representation of the CR model (53 to 70 s). The movie starts with the EM density, contoured at 3.9 σ, of the CR at ~20-Å resolution. Then three-dimensional EM densities of the separately refined core region (average resolution of 3.7 Å, contoured at 5.9 σ) and Nup358 region (average resolution of 4.7 Å, contoured at 3.3 σ) are illustrated at multiple angles. Nucleoporins that are discussed in this paper are color coded. On the basis of these EM maps, a structural model of the CR subunit has been generated.

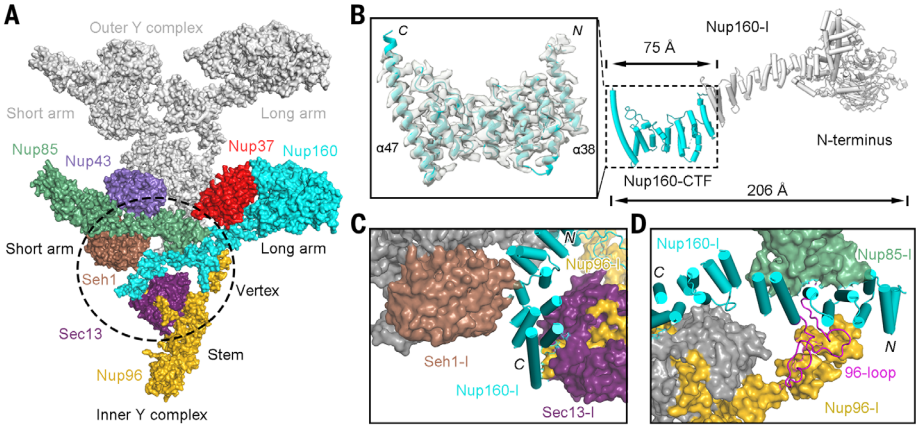


Fig. 2. Nup160-CTF is an organizing center for the vertex of the inner Y complex in the CR. (**A**) Nup160-CTF interacts with several components at the vertex of the inner Y complex. In each Y complex, Nup85, Nup43, and Seh1 constitute the short arm; Nup160 and Nup37 the long arm; and Nup96, Sec13, Nup107, and Nup133 the stem. Shown here is a surface representation for the partial inner and outer Y complexes in the CR. Protein components in the inner and outer Y complexes are denoted with -I and -O, respectively. Components from the inner Y complex are color coded, and the outer Y complex is colored silver. Nup107 and Nup133 are not shown for visual clarity. (**B**) Structure of full-length Nup160, which consists of a seven-bladed β-propeller followed by an extended α-helical domain. Inset: Helices α38 to α47 of Nup160-CTF are well defined in the EM map. The EM density, shown as a semitransparent surface, is contoured at 6 σ for clarity. (**C**) Nup160-I-CTF sandwiched by Seh1-I and Sec13-I. (**D**) The central segments of the Nup160-I α-helical domain are pinched by Nup85-I and Nup96-I. The 96-loop from Nup160, colored magenta, spreads on the surface of Nup96. Except for (B), in which the EM map is shown as a semitransparent surface prepared in ChimeraX, all panels show structures that are presented as cartoon or surface and prepared in PyMol.

featured with well-resolved side chains of a large number of bulky residues (Fig. 4A, figs. S5 and S10).

Nup93-ACE1-O resides at the stem regions of both Y complexes (Fig. 4, A and B). The interfaces between Nup93-ACE1-O and surrounding nucleoporins are discernible in fig. S11. The middle portion, known as the trunk, of Nup93-ACE1-O associates with the trunks of Nup107-O (fig. S11A) and Nup96-I (fig. S11B). The Nup93-Nup96 association appears to be strengthened by Sec13-I, which co-folds with Nup96 and uses its blades to contact the crown of Nup93 (Fig. 4B and fig. S11B). Through these interfaces, Nup93-ACE1-O connects the two Y complexes within one CR subunit.

Nup93-ACE1-I connects Nup205-O with Nup107-I from the neighboring subunit (Fig. 4, A and C). On one end, the trunk of Nup93-ACE1-I associates with the N-terminal domain (NTD) of Nup205-O from the same subunit. On the other end, the tail of Nup93-ACE1-I contacts the lateral side of the tail in Nup107-I from the adjacent subunit (Fig. 4C). Therefore, Nup93-ACE1-I sews two adjacent CR subunits.

Previous biochemical characterization suggested interactions between an N-terminal region of the fungal ortholog of Nup93 and the C-terminal domain (CTD) of Nup205 (*21*). Indeed, helix α5 of Nup93-I may insert into the axial groove of the CTD of Nup205-I, and helix α5 of Nup93-O in one CR subunit (S1) is likely to bind the CTD of Nup205-O in the neighboring subunit (S2) in the same manner (Fig. 4D). The association between Nup93-α5 and Nup205-CTD is almost identical to that observed in the IR and NR subunits (fig. S12) (*22*–*25*).

Five Nup358 clamps in each CR subunit

Five densities each in the shape of a shrimp tail wrap about the stems of the two Y complexes (Fig. 5A). They likely belong to five copies of Nup358, a notion corroborated by our 3.0-Å cryo-EM structure of the middle α-helical region of Nup358 (residues 222 to 738, designated NTD2) (Fig. 5B, figs. S8 and S9). Please refer to the materials and methods for details of the assignment of the five Nup358 molecules, which will be hereafter referred to as clamps 1 to 5.

Clamp 1 through clamp 4 do not interact with each other. They constitute two pairs, clamps 1 and 3 and clamps 2 and 4, which hold the stems of the outer and inner Y complexes, respectively (Fig. 5, C and D, and fig. S13). Clamp 5 connects the two pairs by binding to clamps 1 and 4 (Fig. 5, C and D, and fig. S13A).

The conformational elasticity of Nup358 allows each of these five clamps to adapt to a distinct local environment for optimal interaction with neighboring nucleoporins (figs. S13, B to E, and S14). Aside from the association with the Y complexes (figs. S13, C and D, and S14A), clamps 2 through 5 make distinct contact with Nup93-ACE1-O (Fig. 5, C and D, and fig. S14, B to E). Please refer to the materials and methods for details of the interactions. Clamp 5, the N terminus of which contacts clamp 1 and Nup107-O (fig. S13E), places its C terminus near the central domain of Nup93 (Fig. 5, C and D). Together, these Nup358 molecules encage Nup93-ACE1-O within a cleft between the stems of the two Y complexes (Fig. 5, C and D).

The interaction between Nup358- and ACE1-containing proteins mediates the cross-talk between the Y complexes (fig. S14A). Previous studies identified five candidate interfaces between Y complexes in the CR of vertebrate NPC (*10*–*12*, *18*). Our current study uncovers additional interfaces that are clustered in two regions (fig. S14A and table S5). The first region involves Nup96-I and Nup358 clamp 1,

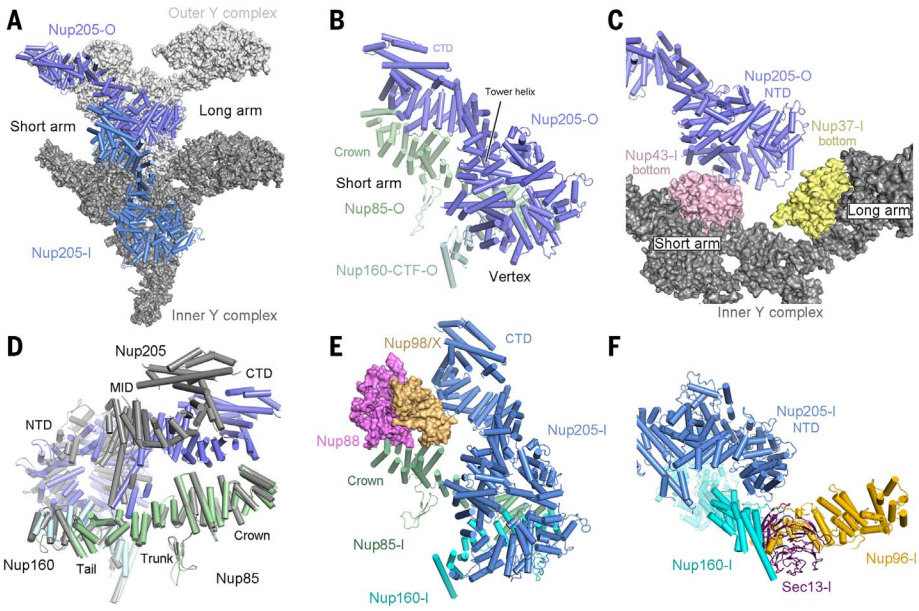


Fig. 3. Nup205 wraps around the Y complexes. (**A**) Nup205-O associates with both the inner and outer Y complexes, whereas Nup205-I only interacts with the inner Y complex. Nup205-O interacts extensively with the short arm and vertex of the outer Y complex, as well as both arms of the inner Y complex. By contrast, Nup205-I only interacts with the short arm and the vertex of the inner Y complex. (**B**) Extensive binding interface between Nup205-O-CTD and Nup85-O. (**C**) Association between Nup205-O and the inner Y complex. (**D**) Structural variations of Nup205. When superimposed relative to Nup-205–associated Nup160 and Nup85, the CTDs of the two Nup205 molecules exhibit pronounced conformational variations. The inner nucleoporins are color coded, and the outer counterparts are colored gray. (**E**) Nup88 and Nup98/X wedge into the space between Nup205-I-CTD and Nup85-I, disrupting their direct association. (**F**) Nup205-I binds to Sec13-I and Nup96-I from the vertex. The structures, shown as cartoon or surface, were prepared in PyMol.

where the tail of Nup96-I contacts the crown of Nup107-O and stacks against the N-terminal α-solenoid of clamp 1 (fig. S14A, middle and right panels). The second one involves Nup205 and the ACE1 of Nup93 molecules, the interfaces of which are described below.

Interfaces between CR subunits

The 3.7-Å EM map of the core region was individually aligned to each of the eight subunits in the 22.2-Å reconstruction, generating a composite model of the CR consisting of the proximal and distal rings (Fig. 6A) (*11*, *12*, *18*). For simplicity, we will refer to two adjacent subunits within the CR as subunit 1 (S1) and subunit 2 (S2) (Fig. 6). Nup205 and Nup93 play critical roles in sewing the neighboring subunits (fig. S15).

The NTD of Nup205-O (S2), together with Nup37-O (S2) and Nup160-O (S2), clench the α-helical domain of Nup133-I (S1) (fig. S15A). The CTD of Nup205-O (S2) associates with the trunk of Nup107-I (S1). The CTD of Nup205-O (S2) harbors the docking site for the outer copy of Nup93-α5, which is likely linked to Nup93-ACE1-O (S1) (fig. S15B). Nup93-ACE1-I (S2) also connects two adjacent subunits. Its tail contacts the lateral side of the tail from Nup107-I (S1) on one side, and its trunk interacts with the NTD of Nup205-O (S2) on the other terminus (fig. S15B).

These structural discoveries establish Nup205 and Nup93 as crucial structural constituents in ring formation.

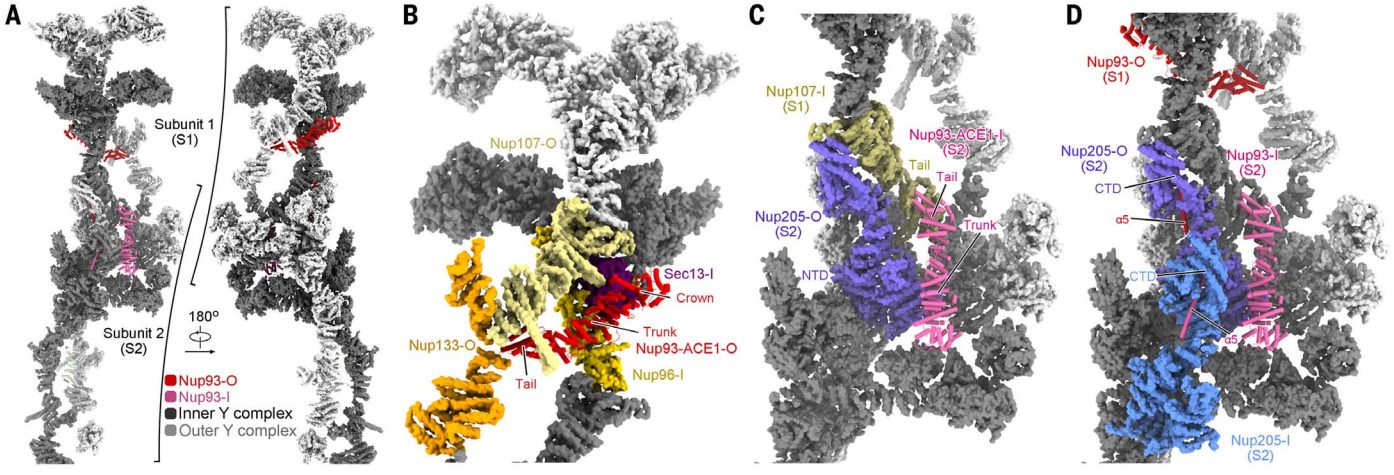


Fig. 4. Nup93 bridges Y complexes and Nup205 molecules. (**A**) Nup93 molecules bridge adjacent Y complexes as well as Nup205 molecules. Identical components from two adjacent CR subunits, designated S1 and S2, are presented. Nup205 and the Y complexes are shown in surface representation. Inner and outer components are colored dark and light gray, respectively. Two copies of Nup93 are present in each CR subunit, with the one closer to the central pore designated as Nup93-I (hot pink) and the other as Nup93-O (red). For visual clarity, only Nup93-O from S1 and Nup93-I from S2 are shown. (**B**) Nup93-ACE1-O connect the stems of the inner and outer Y complexes in each CR subunit. The bridging role is achieved through direct binding to the ACE1 domains in Nup96-I and Nup107-O. These proteins form a triangular ACE1 core at the stem of the Y complexes, which is bolstered by Sec13-I and Nup133-O. Previously defined module names of the ACE1 domain, crown, trunk, and tail are labeled. (**C**) Nup93-ACE1-I bridges two adjacent subunits through direct binding to the ACE1 of Nup107-I from S1 and the NTD of Nup205-O from S2. (**D**) The α5 helix from Nup93 inserts into the axial groove of Nup205-CTD. The structures, shown as cartoon or surface, were prepared in ChimeraX.

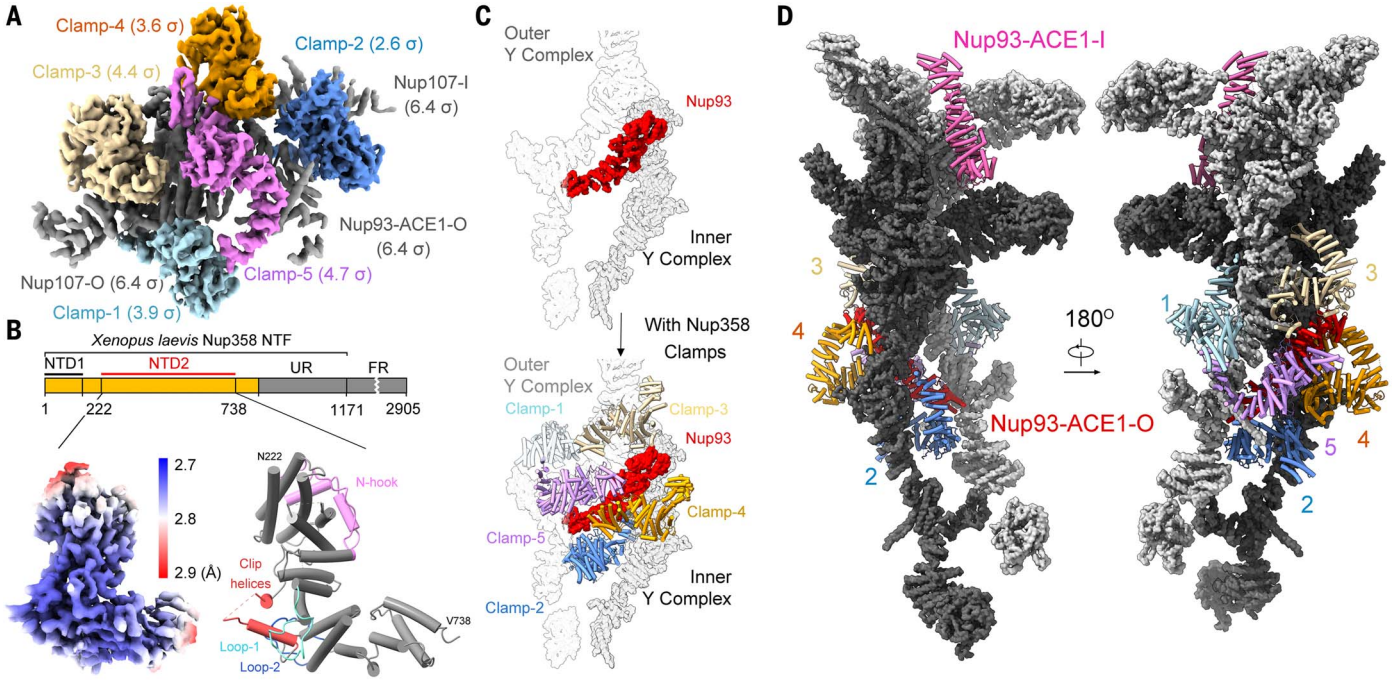


Fig. 5. Five Nup358 clamps connect ACE1 proteins around the stems of the Y complexes. (**A**) EM densities for the five Nup358 clamps within one CR subunit. The EM densities corresponding to the five Nup358 clamps were individually carved out of the refined local map for the Nup358 region using the map segmentation tool in Chimera. The EM maps for the five Nup358 clamps were then individually contoured and colored. The contour level for each clamp is indicated in brackets. In this panel, Nup93 and Nup107 are colored gray. (**B**) Cryo-EM structure of Nup358-NTD2 at 3.0-Å resolution. Left: Resolution map of Nup358-NTD2. The local resolutions might be slightly inflated due to a variety of factors. Please refer to fig. S7C for representative densities. Right: Domain structure of Nup358. UR, unstructured region; FR, functional region. The terminal residues are labeled. Four surface motifs of Nup358-NTD2 (N-hook, loop-1, clip helices, loop-2), which mediate the cross-talk between Nup358-NTD2 and ACE1 proteins within the CR, are color coded. (**C**) Nup93-ACE1-O and Nup358 clamps bridge the two Y complexes. Nup93-ACE1-O binds the stems of both Y complexes (top panel). The five clamps encage the Nup93-ACE1-O (bottom panel). Clamps-4 and -5 directly interact with each other to bridge the inner and outer Y complexes. (**D**) The clamp-1/3 and clamp-2/4 pairs hold the stems of the outer and inner Y complexes, respectively. The overall CR subunit is shown in two opposite views to highlight the relative positions of the five Nup358 clamps and two Nup93 molecules. The inner and outer Y complexes are shown as light- and dark-gray surfaces, respectively. The Nup358 clamps and Nup93 are color coded and highlighted as cylindrical cartoons. Structures in (C) and (D), shown as cartoon or surface, were prepared in ChimeraX.

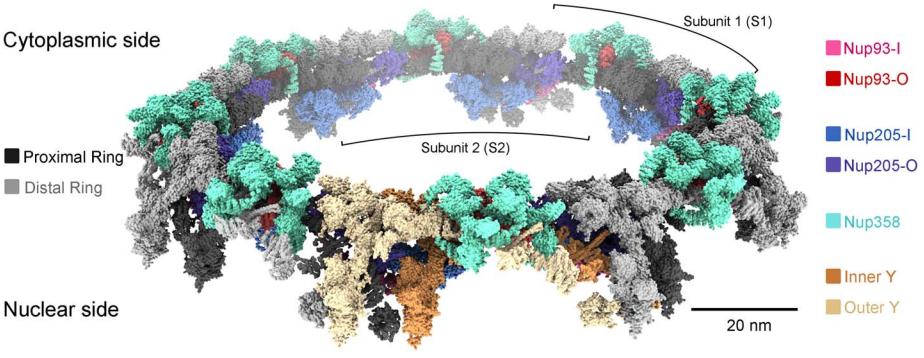


Fig. 6. Structure of the double-layered CR scaffold. Shown is a composite model of the *X. laevis* CR. The model is shown as surface in ChimeraX. In addition to the two concentric Y-complex rings that each comprises eight head-to-tail assembled Y complexes, our present study identifies Nup358, Nup205, and Nup93 to be constituents of the CR scaffold. The inner and outer Y complexes in the subunit closest to reader are colored sandy brown and wheat, respectively, and in the other seven subunits they are colored dark and light gray, respectively.

Discussion

A wealth of structural information at atomic resolution on the NPC have emerged from crystal structures of individual nucleoporins or subcomplexes in the past two decades (*5*, *7*, *26*). Structural studies of the assembled NPC at moderate resolutions have been complemented by other biochemical strategies exemplified by chemical cross-linking and mass spectrometry (*27*–*29*). Collectively, these methods have yielded critical insights into the overall organization of the NPC (*10*–*12*, *30*, *31*). However, it is difficult to obtain accurate distance information, to determine precise stoichiometry, and to pinpoint an interface using these indirect methods. The improved EM map in our present study reveals a number of previously unrecognized features (fig. S5).

The improved model of the CR subunit provides a framework for mechanistic understanding of the NPC assembly. The NPC is subject to assembly and disassembly during a cell cycle. The Y complex has been shown to stay as an intact subcomplex throughout the open mitosis of metazoan (*32*). Our structural finding that the vertebrate-specific Nup160-CTF is at the center of the multiprotein vertex of the Y complex provides a mechanistic explanation for the observed multiple conformations of the short arm in the Y complexes lacking Nup160-CTF (*19*, *33*) (fig. S16). Supporting the functional significance of Nup160-CTF, it was recently reported that residues 1173 to 1436

(corresponding to segments α37 to α47 in the structure) of Nup160 were missing in a patient with steroid-resistant nephrotic syndrome (*34*).

Our EM maps identify two Nup93 molecules within each CR subunit, expanding the list of ACE1-containing nucleoporins (ACE1 proteins) within the CR. In the updated model of the CR subunit, four of seven α-helical domain–containing nucleoporins have the ACE1 fold (*20*), which exhibit a similar conformation in both inner and outer copies (fig. S17). Cross-talk among the ACE1 domains of Nup93-O, Nup107-O, and Nup96-I, which form an ACE1 core at the stem region of the Y complexes, is mediated through interfaces that differ from the canonical crown-crown association (*35*) (Fig. 4B). Our recent analysis of the IR subunit also revealed a homodimerization interface of Nup93 through its crown and tail (*22*). These observations corroborate a key role of the multifaceted ACE1 proteins within the NPC assembly (*20*). Published EM maps (*23*, *36*, *37*) indicate that the Nup93-ACE1-O is also present in the NR subunit from a vertebrate NPC. Together with four molecules of Nup93 in the IR subunit (*12*, *22*, *25*, *30*), there are at least seven Nup93 molecules within each NPC spoke.

Both Nup205 and Nup188 adopt a superhelical fold that is structurally reminiscent of the karyopherin family (*24*, *38*). Our EM map of the core region reveals that two Nup205 molecules, rather than Nup188 molecules, are present within each CR subunit. The IR subunit of the *X. laevis* NPC also comprises two Nup205 molecules (*22*), whereas only one Nup205-O molecule is found in the EM maps of the NR subunit from *X. laevis* (*23*, *36*) and *Homo sapiens* (*37*). These data suggest that at least five Nup205 molecules are present within each NPC spoke. The potentially differential placement of Nup205 in the CR and NR mark one point of structural asymmetry between these two rings.

Among the eight α-helical nucleoporins in the CR subunit, Nup85, Nup96, Nup93, Nup107, and Nup205 each contain a single helical solenoid that confers their structural elasticity (fig. S17). Nup133 and Nup160 each contain an elongated α-helical domain. Unlike the others, Nup358 contains multiple tetratricopeptide repeat (TPR)–like repeats (*39*) that are clustered to two distinct superhelical solenoids. The unique shape and structural features of Nup358 further support our assignment of Nup358 to the five clamp-shaped EM densities. Relative movements of the two solenoids confer additional conformational elasticity on Nup358 (fig. S13B). Identification of the fifth copy of Nup358 expands our understanding of the role of Nup358 and highlights its complex modality in the assembly of CR. Both Nup358 clamp 1 and Nup93-ACE1-O engage in direct interfaces that connect the inner and outer Y complexes (Figs. 4 and 5 and figs. S11 and S14). Such interweaving may strengthen the association between the two concentric rings of the vertebrate CR (*10*). In total, our current model accounts for <30% of the full-length Nup358. The invisible functional domains of Nup358 may project into the cytosol for events such as Ran binding (*7*). Because of residual anisotropy and structural flexibility, the current EM density falls short of sequence assignment for the peripheral regions of the CR subunit, including five Nup358 clamps and the two Nup93 ACE1 domains. Future improvement for these regions of the CR subunit may confer a more conclusive assignment.

In summary, our final model reported in this study, with >200 nucleoporins assigned, is estimated to account for ~90% of the molecular mass for the ordered region of a vertebrate CR. Assembly of the CR scaffold relies on the α-helical nucleoporins; the CR scaffold is structurally solidified by 12 β-propellers in each CR subunit (fig. S18).

Materials and Methods

Cryo-EM sample preparation

The NE from *X. laevis* oocytes was prepared largely as previously described (*9*, *14*) but with the following improvements. First, to reduce conformational heterogeneity of the NPC caused by mechanical distortion, the opened NE was spread onto the grids as gently as possible, with minimal force applied. This practice is in contrast to methods used in previous studies (*9*, *14*), in which mechanical force was applied to ensure flat placement of the NE on the grids. Second, to stabilize the NPC conformation, the NE on the grids was cross-linked for 30 min on ice using 0.5% glutaraldehyde in a low-salt buffer (10 mM Hepes, pH 7.5, 1 mM KCl, and 0.5 mM $MgCl_2$). Third, copper grids were replaced by gold ones in this study. Finally, to improve sample uniformity, each batch of samples was prepared on the same day using oocytes from the same frog. The gold EM grids (R1.2/1.3, R2/1, and R2/2; Quantifoil, Jena, Germany) were blotted for 8 s with a blot force of 15 and vitrified by plunge-freezing into liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific) at 8°C under 100% humidity. The quality of the sample was examined using an FEI Glacios microscope (Thermo Fisher Scientific) operating at 200 kV.

Data acquisition of intact X. laevis NPC

Grids were transferred to a Titan Krios electron microscope (FEI) operating at 300 kV and equipped with a Gatan GIF Quantum energy filter (slit width 20 eV). A total of 46,143 micrographs were recorded with the grids tilting at angles of 0°, 30°, 45°, and 55° (*14*, *40*). A K3 detector (Gatan) was used in the super-resolution mode with a nominal magnification of 64,000×, resulting in a calibrated pixel size of 0.6935 Å for the movie files (fig. S1 and table S1). The movie images were then binned twice during motion correction, arriving at a pixel size of 1.387 Å for the final motion-corrected images. The total dose followed a cosine alpha scheme in which the dose is inversely proportional to the cosine value of the tilting angle. Within each stack, the exposure time for each frame and the dose rate were kept the same. Detailed statistics of data collection are reported in table S1. All frames in each stack were first aligned and summed using MotionCor2 (*41*). Dose weighting was also performed using MotionCor2 (*41*). The average defocus values were set between −1.5 and −3.0 μm and estimated using Gctf (*42*).

Initial model of the CR

We examined all 46,143 micrographs and manually selected 33,747 for further processing. A total of 800,825 particles were manually selected from these micrographs (fig. S2A). Initial defocus estimation was performed as previously described (*14*) before all other data-processing procedures.

We used the 18-Å map of the CR from our previous study (*14*) as the initial reference, which was resampled to a pixel spacing of 11.096 Å and low-pass filtered to 40 Å to reduce potential model bias. We then performed one run of a global search (K = 1) three-dimensional classification, with the solvent mask covering only the CR. The search had 40 iterations. The resulting star files from the last several iterations were then subjected to local search three-dimensional classifications using multiple reference seeds for guidance. All good classes were then merged and duplicated particles were removed, generating a dataset of 660,302 NPC particles. A final round of autorefinement with a solvent mask on the CR resulted in a reconstruction at 22-Å resolution (fig. S2A). The resulting reconstruction was nearly identical to that in our previous study (*14*) (fig. S2B). C8 symmetry was applied throughout this stage of data processing.

Data processing and reconstruction of the CR subunit

We extracted the CR subunit particles on the basis of the alignment parameters of the 22-Å CR reconstruction. We updated the orientation, shift, and defocus parameters for each subunit according to a published protocol (*14*). In short, a cropping center for each subunit was defined in the map, and the Euler angles along with two-dimensional shifts for each subunit particle were deduced from the CR particle. The recentering procedure was repeated eight times, defining a three-dimensional geometric relationship among eight subunits of the same CR. A total of 5,148,474 particles of the CR subunit were extracted using a box size of 256 and a binned pixel size of 2.774 Å (fig. S3A). We then performed three rounds of CTF refinement with

geometrical restraint among subunits of the same CR and five rounds of guided multireference three-dimensional classification. This practice allowed the selection of 2,477,433 CR subunits, which yielded a reconstruction of the core region at 5.77-Å resolution. To fully use the dataset, these subunits were projected back to the original CR particles. The defocus values of all subunit particles within the same CR were pooled together to calculate a corrected average defocus value for the center of mass of each CR particle. All subunits of this CR were then reextracted using the updated defocus value deduced from the corrected average defocus of this CR particle. Through this procedure, 4,528,642 subunit particles from 581,882 CR particles were extracted, resulting in a reconstruction at 5.6 Å after autorefinement (fig. S3A).

This dataset was then used for reextraction with a box size of 400 and pixel size of 1.387 Å (fig. S3B). Following our published protocol (*14*), data processing beyond this point was performed individually for four datasets defined by the tilting angles: tilt0, tilt30, tilt45, and tilt55. This was because of our observation that reconstruction at relatively high resolution is easily biased toward low-tilt datasets due to their relatively high signal-to-noise ratio. In fact, the high-tilt and low-tilt datasets exhibited quite different motion statistics (fig. S1). The four datasets were then separately reconstructed to obtain an estimation of signal-to-noise ratio for their respective tilting angles. After polishing and CTF refinements, the four datasets were pooled for three-dimensional autorefinements and three-dimensional classifications.

A second strategy to enhance the quality of the reconstruction is to improve the results of particle polishing. In particle polishing, initial movement tracks are obtained either through global frame alignment or polynomial motion models determined from local motion trajectories (*43*). The magnitude of the sample drifts in the tilted datasets appears much greater than that in the untilted ones (fig. S1). This problem, along with increased sample thickness at high-tilt angles, prevents the local motion model from accommodating sample deformation. An initial motion model estimation using MotionCor2 resulted in 18,619 failed attempts to explain local sample deformation in 38,583 movie stacks using the polynomial model. In particular, sample deformation was successfully explained for only 6825 movie stacks out of a total of 22,132 for the tilt45 and tilt55 datasets. Errors introduced by the polynomial motion model undermine the accuracy of the initial movement tracks for particle polishing, thus affecting the final outcome. To address this issue, we took advantage of the large size of a single NPC particle and attempted to estimate only the local motion around each NPC particle instead of dividing the micrograph into rectangular patches and estimating motion for each patch. A major rationale for this practice is that empty patches, which occur often, may engender misalignment, thereby compromising the accuracy of the polynomial motion model. To improve alignment, we used a polynomial motion model similar to those implemented in MotionCor2 (*41*), RELION (*44*), and WARP (*45*) to regularize movement tracks of particles that belong to the same micrograph. This polynomial model was iteratively refined until convergence. A final round of patch-based alignment using only information around each NPC particle was performed to account for any residual movement of each particle. Finally, any micrograph that failed to converge or had residual motion of >15 pixels in any direction was removed.

Using the above strategy, we were able to rescue 13,758 micrographs from 18,619 that failed to provide reliable local motion models. We separately reconstructed the four tilt datasets and performed one round of CTF refinement. The results were directly subjected to autorefinement, resulting in a reconstruction of the core region at 4.5-Å resolution (fig. S3B). The combined dataset was then redivided into four subsets according to tilting angle and subjected to particle polishing and CTF refinement in RELION. The resulting particles were again pooled together and subjected to autorefinement and multireference three-dimensional classifications. The above cycle was repeated several times. In the last few cycles, magnification anisotropy refinement and per-micrograph astigmatism refinement were also performed to reduce distortion due to incorrect magnification and astigmatism. Additionally, the box size of the used particles was enlarged during particle polishing for the last few cycles using the window and scale option from relion_motion_refine. These measures resulted in a final average resolution of 3.7 Å from 1,279,270 particles (fig. S3B and table S1).

For the Nup358 region, the particles were subjected to local refinement in cryoSparc (*46*) to reach an average resolution of 4.7 Å (fig. S3B, and table S1). The EM maps in the core region display distinct features for secondary structural elements and some of the bulky amino acid side chains (figs. S4 and S5). These features facilitated sequence assignment of the nucleoporins and identification of protein-protein interfaces.

Expression and purification of Nup358-NTF

The cDNA encoding the N-terminal 1300 amino acids of *X. laevis* Nup358 (Uniprot: A0A1L8HGL2) was synthesized with codon optimization (Qinglan Biotech). The N-terminal fragment (residues 1 to 1171, referred to as Nup358-NTF) with an N-terminal Flag tag was cloned into the pCAG vector. This construct was verified by DNA sequencing. HEK293F cells (Invitrogen) were cultured in SMM 293T-II medium (SinoBiological) supplemented with 5% CO_2 in a Multitron-Pro shaker at 130 rpm at 37°C. The cells were transfected at a density of ~2 × 10^6 cells/ml. SMS 293-SUPI (SinoBiological) was supplemented into the culture 24 hours after the transfection. The cells were then cultured for additional 24 to 36 hours.

The transfected HEK293F cells were harvested by centrifugation at 3800*g* and resuspended in a lysis buffer containing 25 mM Tris-HCl, pH 8.0, 150 mM NaCl, and a cocktail of protease inhibitors (VWR). The final concentrations of the inhibitors were 2 mM for phenylmethylsulfonyl fluoride (PMSF), 5.2 μg/ml for aprotinin, 2.8 μg/ml for pepstatin, and 10 μg/ml for leupeptin. The cells were lysed by ultrasonication (Vibra-Cell, SONICS). After centrifugation at 30,000*g* for 1 hour, the supernatant was loaded into an anti-Flag M2 affinity gel (Sigma-Aldrich) column and eluted with the lysis buffer supplemented with 200 μg/ml Flag peptide. The eluted fraction was concentrated using a 50-kDa cut-off Centricon (Millipore) and applied to size-exclusion chromatography (Superose-6, GE Healthcare). The peak fractions were analyzed by SDS-polyacrylamide gel electrophoresis (fig. S6).

Cryo-EM data acquisition for Nup358-NTF

An aliquot of 4 μl of freshly purified Nup358-NTF at a concentration of ~2 mg/ml was placed on glow-discharged holey carbon grids (Quantifoil Au 300 mesh, R1.2/1.3). Grids were blotted for 3.5 s and plunge-frozen in liquid ethane cooled by liquid nitrogen using Vitrobot Mark IV (Thermo Fisher) at 8°C under 100% humidity. The grids were transferred to a Titan Krios electron microscope (Thermo Fisher) operating at 300 kV and equipped with a GIF Quantum energy filter (Gatan). Using the same setup as that for intact *X. laevis* NPC data acquisition, a total of 17,147 movie stacks were recorded. WARP was used for all subsequent preprocessing, including CTF estimation and motion correction (*45*).

Image processing for Nup358-NTF

Out of a subset of 1176 micrographs, 627,188 particles were first autopicked using Topaz (*47*), with its default model on micrographs prebinned by eight times. Particles were extracted using a box size of 100 and a pixel size of 2.774 Å. These particles were then subjected to four rounds of two-dimensional classifications in cryoSparc (*46*), yielding 201,480 particles. The good particles were then converted to RELION-compatible star files using the csparc2star.py program from the pyem suite (*48*). These particles were used to train a deep learning model in Topaz. Relying on this model, we picked 2,418,968 particles from 17,147 micrographs using a box size of 100 and a binned

pixel size of 2.774 Å. These particles were directly subjected to two additional rounds of two-dimensional classifications in RELION, yielding a dataset containing 2,047,018 particles (fig. S7A). These particles were then reextracted using a box size of 200 and a pixel size of 1.387 Å and imported into cryoSparc. Eight repetitions of initial model generation were executed, and the best initial model was selected on the basis of visual inspection.

We then used csparc2star.py to convert the 213,534 particles assigned to this class into RELION-compatible star files and used the orientational parameters contained within to generate a RELION-compatible reconstruction. Using this initial model, one round of random-phase three-dimensional classification with the random_mask and solvent_mask options turned on was performed on the whole dataset. Similar to a previously published procedure (*49*), all other references except for class 1 were phase randomized to generate a "bad" reference with the user-specified phase randomization upper and lower limits. To further distinguish protein from solvent, the EM density outside of the user-supplied random_mask was kept unchanged in the "bad" references; the EM density inside the random_mask was subjected to the phase randomization procedure. Conversely, the EM density inside the random_mask was kept unchanged in the "good" reference and that outside was subjected to phase randomization.

Using this approach, we were able to perform image alignment for the relatively small Nup358-NTF protein for up to 40 iterations. After the global search random phase classification procedure, data stars from the last several iterations from the global search run were subjected to local search multireference three-dimensional classifications. The good classes were merged and duplicated particles were removed. Finally, the remaining 744,385 particles were imported into cryoSparc for a final round of local refinement to yield reconstructions at an average resolution of 3.0 Å for two separate regions that were flexible relative to each other (fig. S7A and table S1). The final EM reconstruction exhibits clear features for sequence assignment (fig. S7, B and C).

Structural modeling of the CR subunit

The 4.7-Å map of the Nup358 region contains five similar, clamp-shaped density patches. To facilitate modeling of these five clamps, we first generated the atomic model and assigned sequence register for Nup358-NTD2 (residues 222 to 738) on the basis of the 3.0-Å EM map of Nup358-NTF (fig. S8). The atomic model of Nup358-NTD2 was refined against the 3.0-Å map using Phenix real-space refine (*50*) with secondary structure restraints.

The final atomic model of Nup358-NTD2 can be placed into each of these five density patches with little adjustment (fig. S9). On the basis of the x-ray structure of human Nup358-NTD1 (*39*), we generated a homology model for NTD1 (residues 1 to 145) of Nup358 from *X. laevis*. Nup358-NTD1 and Nup358-NTD2 were separately docked into each of the five clamp-shaped density areas. In each of these five areas, this practice leaves a similar density patch unfilled between NTD1 and NTD2. This unfilled density patch was assigned to the sequences between NTD1 and NTD2 (residues 146 to 221), which were predicted to be α-helical. We generated a model for residues 1 to 738 of Nup358, which was individually fit into each of the five clamp-shaped density areas. This observation suggests the presence of a fifth Nup358 molecule in addition to four speculated Nup358 molecules in the Nup358 region (fig. S9) (*29*).

Apart from the newly resolved structure of Nup358-NTD2, modeling for the CR subunit were largely based on the coordinates of *X. laevis* CR from our previous study [Protein Data Bank (PDB) ID: 6LK8] (*14*), which were fitted into the new reconstruction of the CR subunit using Chimera (*51*). For the core region, the secondary structural elements were manually checked and adjusted on the basis of the much-improved EM density map using Coot (*52*). The reconstruction of the core region at 3.7-Å resolution allowed us to identify a large number of bulky residues in components of both the inner and outer Y complexes and in Nup205. We de novo modeled 10 α-helices in the newly assigned C-terminal fragment of Nup160 (Nup160-CTF), which was manually adjusted on the basis of the sequence conservation, and secondary structural elements were extracted from the AlphaFold model (*15*). For the less-well-resolved N-terminal regions of Nup160-O, modeling was performed by docking of the coordinates of the N-terminal regions of Nup160-I into the EM density of core region low-pass filtered to 10 Å, in which the outer long arm is clearly discernible. In addition, the β-propeller of Nup88, the autoproteolytic domain of Nup98, and the CTD of Nup155 were modeled using the predicted structures. The current EM density does not allow for reliable docking for Nup98. We refer to this model as Nup98/X herein.

For modeling of Nup205 molecules, we generated the secondary structural elements of Nup205 on the basis of structural alignment of its functional orthologs and the crystal structure of Nup205 orthologs in fungi using the AlphaFold model (*15*). Relying on structural features of the IR subunit in various organisms (*22*, *24*, *25*), previous biochemical characterizations of the Nup205 ortholog in fungi (*21*), and the EM density (fig. S5), we also modeled two Nup93 α5 helices inserted into the CTD of Nup205 molecules. Our model covers the entire α-helical region of Nup205, including a CTD- and vertebrate-specific TAIL-C (*14*) that harbor the docking site of Nup93 α5.

In the center of the Nup358 region, the bridge domain of unknown identity connects the Nup358 clamps and associates with other surrounding nucleoporins (*14*). In both the NR from frog and human NPC, a rod of EM density with nearly identical shape resides in the same place as the bridge domain within the two concentric Y complex rings. Previous analysis pointed to TPR, a nuclear basket–specific nucleoporin, as a candidate that may occupy this location (*10*). The 4.7-Å EM map of the Nup358 region allowed generation of a poly-Ala model for the bridge domain. The poly-Ala model was used to search the PDB using the DALI server (*53*). This search led to identification of a number of potential candidates, of which the yeast NPC component Nic96 tops the list, suggesting that the bridge domain is Nup93, the *X. laevis* ortholog of Nic96.

Near the end of our study, DeepMind released a database of 20,000 predicted structures of the human proteome, which includes that of human Nup93 (*54*). To our satisfaction, the poly-Ala model of the bridge domain and the predicted model of human Nup93 display a root mean square deviation of 2.43 Å over 416 aligned Cα atoms (fig. S10A). In fact, the predicted structure of human Nup93 can be directly docked into the EM density map for the bridge domain without adjustment (fig. S10B). The nearly perfect fitting of human Nup93 at the secondary structure level strongly supports the identification of the bridge domain to be Nup93. This conclusion is further supported by mass spectrometric analysis of the cross-linked *X. laevis* NE, in which Nup93 was found to be mainly cross-linked to Nup358 among the nucleoporins (fig. S10C).

Using AlphaFold (*15*), we generated a predicted atomic model for *X. laevis* Nup93, which fits the EM density map exceedingly well (fig. S10D). The surface loop region of the predicted model for *X. laevis* Nup93 was manually adjusted into the EM density. This final model of *X. laevis* Nup93 is almost identical to our initial poly-Ala model of the bridge domain but contains more features. The featured model of *X. laevis* Nup93 (fig. S10E) was used for all subsequent structural analysis. The structure of *X. laevis* Nup93 α-solenoid (residues 180 to 820), with a fold-back conformation characteristic of ACE1 (*20*), consists of a trunk module (α6 to α9 and α18 to α28), a crown module (α10 to α17), and a tail module (α29 to α37) (Fig. 4B and fig. S10E). In addition, in the N terminus of the Nup93 α-solenoid, an extended α-helix (α5) inserted in the CTD of Nup205 was also modeled (figs. S5 and S10E). The corresponding region in yeast Nic96 was reported to bind the CTD of Nup192 (the fungal homolog of Nup205) (*21*). Secondary structural elements were assigned on the basis of the sequence alignment of

functional orthologs of Nup93 and the AlphaFold model. The EM maps also allow identification of protein-protein interfaces mediated by Nup93 (figs. S11, S12, and S14).

With one Nup93-ACE1 assigned to the bridge domain and two Nup93 α5 helices identified to be inserted into the CTD of Nup205 molecules, we investigated whether a second copy of Nup93-ACE1 was present within the CR subunit. The current molecular model left a chunk of unassigned density spanning the Nup107-I from one subunit and Nup205-O from the adjacent subunit that adopts a similar shape as those of the bridge domain. Local refinements focusing on this chunk of density revealed that a second copy of Nup93-ACE1 was indeed present in the CR subunit. Direct docking of the model of Nup93-ACE1-O into this region resulted in a nearly perfect match (fig. S10F).

Modeling and structural analyses of other nucleoporins were facilitated by sequence alignment (*55*, *56*) of functional orthologs of the structurally defined CR components and AlphaFold-predicted models. Overall, the final model of the CR subunit includes the inner and outer Y complexes, two Nup205 molecules, five Nup358 clamps, two Nup93 molecules, and one copy each of Nup88, Nup98/X, and Nup155. This model includes 19,037 amino acids in 749 α-helices and 380 β-strands (tables S2 and S3). The final model was refined using Phenix (*50*) with secondary structure restraints and validated through examination of the Molprobity scores and statistics of the Ramachandran plots (*57*) (table S1). Together with the EM density map, this model allows structural analysis of Nup205 (Fig. 3), Nup93 (Fig. 4 and figs. S10, S11, S12, and S14), and Nup358 (Fig. 5 and figs. S13 and S14) clamps. In addition, the final model provides a basis for the identification of inter-subunit protein-protein interfaces (fig. S15) and structural comparisons among Y complexes from various species (fig. S16) and between individual components of the CR (fig. S17 and table S4).

The figures for structures and maps were generated using Pymol, Chimera, or ChimeraX (*51*, *58*). Sequence alignments were done in Clustal Omega (*55*) and ESPript 3.0 (*56*).

REFERENCES AND NOTES

1. H. B. Schmidt, D. Görlich, Transport selectivity of nuclear pores, phase separation, and membraneless organelles. *Trends Biochem. Sci.* **41**, 46–61 (2016). doi: 10.1016/j.tibs.2015.11.001; pmid: 26705895
2. R. Ungricht, U. Kutay, Mechanisms and functions of nuclear envelope remodelling. *Nat. Rev. Mol. Cell Biol.* **18**, 229–245 (2017). doi: 10.1038/nrm.2016.153; pmid: 28120913
3. P. De Magistris, W. Antonin, The dynamic nature of the nuclear envelope. *Curr. Biol.* **28**, R487–R497 (2018). doi: 10.1016/j.cub.2018.01.073; pmid: 29689232
4. J. Fernandez-Martinez, M. P. Rout, A jumbo problem: Mapping the structure and functions of the nuclear pore complex. *Curr. Opin. Cell Biol.* **24**, 92–99 (2012). doi: 10.1016/j.ceb.2011.12.013; pmid: 22321828
5. T. U. Schwartz, The structure inventory of the nuclear pore complex. *J. Mol. Biol.* **428** (10 Pt A), 1986–2000 (2016). doi: 10.1016/j.jmb.2016.03.015; pmid: 27016207
6. M. Beck, E. Hurt, The nuclear pore complex: Understanding its function through structural insight. *Nat. Rev. Mol. Cell Biol.* **18**, 73–89 (2017). doi: 10.1038/nrm.2016.147; pmid: 27999437
7. D. H. Lin, A. Hoelz, The structure of the nuclear pore complex (an update). *Annu. Rev. Biochem.* **88**, 725–783 (2019). doi: 10.1146/annurev-biochem-062917-011901; pmid: 30883195
8. J. Fernandez-Martinez, M. P. Rout, One ring to rule them all? Structural and functional diversity in the nuclear pore complex. *Trends Biochem. Sci.* **46**, 595–607 (2021). doi: 10.1016/j.tibs.2021.01.003; pmid: 33563541
9. Y. Zhang *et al.*, Molecular architecture of the luminal ring of the *Xenopus laevis* nuclear pore complex. *Cell Res.* **30**, 532–540 (2020). doi: 10.1038/s41422-020-0320-y; pmid: 32367042
10. A. von Appen *et al.*, In situ structural analysis of the human nuclear pore complex. *Nature* **526**, 140–143 (2015). doi: 10.1038/nature15381; pmid: 26416747
11. K. H. Bui *et al.*, Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* **155**, 1233–1243 (2013). doi: 10.1016/j.cell.2013.10.055; pmid: 24315095
12. D. H. Lin *et al.*, Architecture of the symmetric core of the nuclear pore. *Science* **352**, aaf1015 (2016). doi: 10.1126/science.aaf1015; pmid: 27081075
13. M. Lutzmann, R. Kunze, A. Buerer, U. Aebi, E. Hurt, Modular self-assembly of a Y-shaped multiprotein complex from seven nucleoporins. *EMBO J.* **21**, 387–397 (2002). doi: 10.1093/emboj/21.3.387; pmid: 11823431
14. G. Huang *et al.*, Structure of the cytoplasmic ring of the *Xenopus laevis* nuclear pore complex by cryo-electron microscopy single particle analysis. *Cell Res.* **30**, 520–531 (2020). doi: 10.1038/s41422-020-0319-4; pmid: 32376910
15. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). doi: 10.1038/s41586-021-03819-2; pmid: 34265844
16. T. Boehmer, S. Jeudy, I. C. Berke, T. U. Schwartz, Structural and functional studies of Nup107/Nup133 interaction and its implications for the architecture of the nuclear pore complex. *Mol. Cell* **30**, 721–731 (2008). doi: 10.1016/j.molcel.2008.04.022; pmid: 18570875
17. M. Kampmann, G. Blobel, Three-dimensional structure and flexibility of a membrane-coating module of the nuclear pore complex. *Nat. Struct. Mol. Biol.* **16**, 782–788 (2009). doi: 10.1038/nsmb.1618; pmid: 19503077
18. T. Stuwe *et al.*, Nuclear pores. Architecture of the nuclear pore complex coat. *Science* **347**, 1148–1152 (2015). doi: 10.1126/science.aaa4136; pmid: 25745173
19. K. Kelley, K. E. Knockenhauer, G. Kabachinski, T. U. Schwartz, Atomic structure of the Y complex of the nuclear pore. *Nat. Struct. Mol. Biol.* **22**, 425–431 (2015). doi: 10.1038/nsmb.2998; pmid: 25822992
20. S. G. Brohawn, N. C. Leksa, E. D. Spear, K. R. Rajashankar, T. U. Schwartz, Structural evidence for common ancestry of the nuclear pore complex and vesicle coats. *Science* **322**, 1369–1373 (2008). doi: 10.1126/science.1165886; pmid: 18974315
21. T. Stuwe *et al.*, Architecture of the fungal nuclear pore inner ring complex. *Science* **350**, 56–64 (2015). doi: 10.1126/science.aac9176; pmid: 26316600
22. G. Huang *et al.*, Cryo-EM structure of the inner ring from the *Xenopus laevis* nuclear pore complex. *Cell Res.* **32**, 451–460 (2022). doi: 10.1038/s41422-022-00633-x; pmid: 35301439
23. G. Huang *et al.*, Cryo-EM structure of the nuclear ring from *Xenopus laevis* nuclear pore complex. *Cell Res.* **32**, 349–358 (2022). doi: 10.1038/s41422-021-00610-w; pmid: 35177819
24. C. W. Akey *et al.*, Comprehensive structure and functional adaptations of the yeast nuclear pore complex. *Cell* (2021). pmid: 34982960
25. Z. Li *et al.*, Near-atomic structure of the inner ring of the *Saccharomyces cerevisiae* nuclear pore complex. *Cell Res.* **32**, 437–450 (2022). doi: 10.1038/s41422-022-00632-y; pmid: 35301440
26. A. Hoelz, E. W. Debler, G. Blobel, The structure of the nuclear pore complex. *Annu. Rev. Biochem.* **80**, 613–643 (2011). doi: 10.1146/annurev-biochem-060109-151030; pmid: 21495847
27. M. P. Rout *et al.*, The yeast nuclear pore complex: Composition, architecture, and transport mechanism. *J. Cell Biol.* **148**, 635–652 (2000). doi: 10.1083/jcb.148.4.635; pmid: 10684247
28. J. M. Cronshaw, A. N. Krutchinsky, W. Zhang, B. T. Chait, M. J. Matunis, Proteomic analysis of the mammalian nuclear pore complex. *J. Cell Biol.* **158**, 915–927 (2002). doi: 10.1083/jcb.200206106; pmid: 12196509
29. A. Ori *et al.*, Cell type-specific nuclear pores: A case in point for context-dependent stoichiometry of molecular machines. *Mol. Syst. Biol.* **9**, 648 (2013). doi: 10.1038/msb.2013.4; pmid: 23511206
30. J. Kosinski *et al.*, Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352**, 363–365 (2016). doi: 10.1126/science.aaf0643; pmid: 27081072
31. S. J. Kim *et al.*, Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475–482 (2018). doi: 10.1038/nature26003; pmid: 29539637
32. J. S. Glavy *et al.*, Cell-cycle-dependent phosphorylation of the nuclear pore Nup107-160 subcomplex. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3811–3816 (2007). doi: 10.1073/pnas.0700058104; pmid: 17360435
33. M. Allegretti *et al.*, In-cell architecture of the nuclear pore and snapshots of its turnover. *Nature* **586**, 796–800 (2020). doi: 10.1038/s41586-020-2670-5; pmid: 32879490
34. F. Zhao *et al.*, Mutations in NUP160 are implicated in steroid-resistant nephrotic syndrome. *J. Am. Soc. Nephrol.* **30**, 840–853 (2019). doi: 10.1681/ASN.2018080786; pmid: 30910934
35. V. Nagy *et al.*, Structure of a trimeric nucleoporin complex reveals alternate oligomerization states. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 17693–17698 (2009). doi: 10.1073/pnas.0909373106; pmid: 19805193
36. L. Tai *et al.*, 8 Å structure of the outer rings of the *Xenopus laevis* nuclear pore complex obtained by cryo-EM and AI. *Protein Cell* (2022). doi: 10.1007/s13238-021-00895-y; pmid: 35015240
37. S. Mosalaganti *et al.*, Artificial intelligence reveals nuclear pore complexity. bioRxiv 465776 [Preprint] (2021). doi: 10.1101/2021.10.26.465776
38. K. R. Andersen *et al.*, Scaffold nucleoporins Nup188 and Nup192 share structural and functional properties with nuclear transport receptors. *eLife* **2**, e00745 (2013). doi: 10.7554/eLife.00745; pmid: 23795296
39. S. A. Kassube *et al.*, Crystal structure of the N-terminal domain of Nup358/RanBP2. *J. Mol. Biol.* **423**, 752–765 (2012). doi: 10.1016/j.jmb.2012.08.026; pmid: 22959972
40. Y. Z. Tan *et al.*, Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017). doi: 10.1038/nmeth.4347; pmid: 28671674
41. 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
42. 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
43. J. Zivanov, T. Nakane, S. H. W. Scheres, A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019). doi: 10.1107/S205225251801463X; pmid: 30713699
44. 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
45. D. Tegunov, P. Cramer, Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019). doi: 10.1038/s41592-019-0580-y; pmid: 31591575
46. 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
47. T. Bepler *et al.*, Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019). doi: 10.1038/s41592-019-0575-8; pmid: 31591578
48. D. Asarnow, E. Palovcak, Y. Cheng, "UCSF pyem v0.5," Zenodo (2019); doi: 10.5281/zenodo.3576630
49. X. Gong *et al.*, Structural Insights into the Niemann-Pick C1 (NPC1)-Mediated Cholesterol Transfer and Ebola Infection. *Cell* **165**, 1467–1478 (2016). doi: 10.1016/j.cell.2016.05.022; pmid: 27238017
50. P. D. Adams *et al.*, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010). doi: 10.1107/S0907444909052925; pmid: 20124702
51. E. F. Pettersen *et al.*, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254

52. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004). doi: 10.1107/S0907444904019158; pmid: 15572765
53. L. Holm, Using Dali for protein structure comparison. *Methods Mol. Biol.* **2112**, 29–42 (2020). doi: 10.1007/978-1-0716-0270-6_3; pmid: 32006276
54. K. Tunyasuvunakool *et al*., Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021). doi: 10.1038/s41586-021-03828-1; pmid: 34293799
55. F. Madeira *et al*., The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019). doi: 10.1093/nar/gkz268; pmid: 30976793
56. X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014). doi: 10.1093/nar/gku316; pmid: 24753421
57. C. J. Williams *et al*., MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018). doi: 10.1002/pro.3330; pmid: 29067766
58. T. D. Goddard *et al*., UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018). doi: 10.1002/pro.3235; pmid: 28710774
59. S. H. 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

ACKNOWLEDGMENTS

We thank the staff of the Cryo-EM Core and the Computing Core of Westlake University for technical support. **Funding:** This work was supported by the National Natural Science Foundation of China (grant 31930059 to Y.S.), Westlake University (start-up funds to Y.S.), The China Postdoctoral Science Foundation (grant 2021M692888 to X. Zhan), and the National Postdoctoral Program for Innovative Talents of China (grant BX2021268 to X. Zhan).
Author contributions: Y.S. conceived and supervised the project. X. Zhu, P.W., and C.Z. prepared the sample. G.H., C.Z., X. Zhan, K.L., and J.L. collected the EM data. G.H., C.Z., X. Zhan, and K.L. processed the EM data. G.H. performed the cryo-EM SPA calculation. X. Zhan built the models. Q.Z., C.Y., Q.T., and M.L. provided critical support. All authors analyzed the structure. G.H., C.Z., X. Zhan, and Y.S. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The atomic coordinates of the CR subunit and NTD2 of Nup358 have been deposited in the PDB with accession codes 7FIK and 7FIL, respectively. The EM maps for the CR core region, the Nup358 region, and the NTD2 for Nup358 have been deposited in the Electron Microscopy Data Bank with the accession codes EMD-31600, EMD-31601, and EMD-31602, respectively. **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.abl8280
Figs. S1 to S18
Tables S1 to S5
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 7 August 2021; resubmitted 30 September 2021
Accepted 10 May 2022
10.1126/science.abl8280

RESEARCH ARTICLE SUMMARY

NUCLEAR PORE COMPLEX

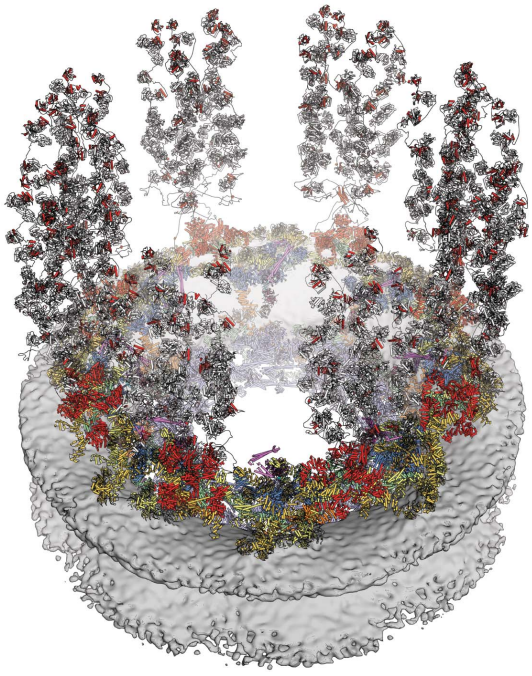
Architecture of the cytoplasmic face of the nuclear pore

Christopher J. Bley†, Si Nie†, George W. Mobbs†, Stefan Petrovic†, Anna T. Gres†, Xiaoyu Liu†, Somnath Mukherjee, Sho Harvey, Ferdinand M. Huber, Daniel H. Lin, Bonnie Brown, Aaron W. Tang, Emily J. Rundlet, Ana R. Correia, Shane Chen, Saroj G. Regmi, Taylor A. Stevens, Claudia A. Jette, Mary Dasso, Alina Patke, Alexander F. Palazzo, Anthony A. Kossiakoff, André Hoelz*

INTRODUCTION: The subcellular compartmentalization of eukaryotic cells requires selective transport of folded proteins and protein–nucleic acid complexes. Embedded in nuclear envelope pores, which are generated by the circumscribed fusion of the inner and outer nuclear membranes, nuclear pore complexes (NPCs) are the sole bidirectional gateways for nucleocytoplasmic transport. The ~110-MDa human NPC is an ~1000-protein assembly that comprises multiple copies of ~34 different proteins, collectively termed nucleoporins. The symmetric core of the NPC is composed of an inner ring encircling the central transport channel and outer rings formed by Y-shaped coat nucleoporin complexes (CNCs) anchored atop both sides of the nuclear envelope. The outer rings are decorated with compartment-specific asymmetric nuclear basket and cytoplasmic filament nucleoporins, which establish transport directionality and provide docking sites for transport factors and the small guanosine triphosphatase Ran. The cytoplasmic filament nucleoporins also play an essential role in the irreversible remodeling of messenger ribonucleoprotein particles (mRNPs) as they exit the central transport channel. Unsurprisingly, the NPC's cytoplasmic face represents a hotspot for disease-associated mutations and is commonly targeted by viral virulence factors.

RATIONALE: Previous studies established a near-atomic composite structure of the human NPC's symmetric core by combining (i) biochemical reconstitution to elucidate the interaction network between symmetric nucleoporins, (ii) crystal and single-particle cryo–electron microscopy structure determination of nucleoporins and nucleoporin complexes to reveal their three-dimensional shape and the molecular details of their interactions, (iii) quantitative docking in cryo–electron tomography (cryo-ET) maps of the intact human NPC to uncover nucleoporin stoichiometry and positioning, and (iv) cell-based assays to validate the physiological relevance of the biochemical and structural findings. In this work, we extended our approach to the cytoplasmic filament nucleoporins to reveal the near-atomic architecture of the cytoplasmic face of the human NPC.

RESULTS: Using biochemical reconstitution, we elucidated the protein-protein and protein-RNA interaction networks of the human and *Chaetomium thermophilum* cytoplasmic filament nucleoporins, establishing an evolutionarily conserved heterohexameric cytoplasmic filament nucleoporin complex (CFNC) held together by a central heterotrimeric coiled-coil hub that tethers two separate mRNP-remodeling complexes. Further biochemical analysis and determination of a series of crystal structures revealed that the metazoan-specific cytoplasmic filament nucleoporin NUP358 is composed of 16 distinct domains, including an N-terminal S-shaped α-helical solenoid followed by a coiled-coil oligomerization element, numerous Ran-interacting domains, an E3 ligase domain, and a C-terminal prolyl-isomerase domain. Physiologically validated quantitative docking into cryo-ET maps of the intact human NPC revealed that pentameric NUP358 bundles, conjoined by the oligomerization element, are anchored through their N-terminal domains to the central stalk regions of the CNC, projecting flexibly attached domains as far as ~600 Å into the cytoplasm. Using cell-based assays, we demonstrated that NUP358 is dispensable for the architectural integrity of the assembled interphase NPC and RNA export but is required for efficient translation. After NUP358 assignment, the remaining 4-shaped cryo-ET density matched the dimensions of the CFNC coiled-coil hub, in close proximity to an outer-ring NUP93. Whereas the N-terminal NUP93 assembly sensor motif anchors the properly assembled related coiled-coil channel nucleoporin heterotrimer to the inner ring, biochemical reconstitution confirmed that the NUP93 assembly sensor is reused in anchoring the CFNC to the cytoplasmic face of the human NPC. By contrast, two *C. thermophilum* CFNCs are anchored by a divergent mechanism that involves assembly sensors located in unstructured portions of two CNC nucleoporins. Whereas unassigned cryo-ET density occupies the NUP358 and CFNC binding sites on the nuclear face, docking of the nuclear basket component ELYS established that the equivalent position on the cytoplasmic face is unoccupied, suggesting that mechanisms other than steric competition promote asymmetric distribution of nucleoporins.



Cytoplasmic face of the human NPC. Near-atomic composite structure of the NPC generated by docking high-resolution crystal structures into a cryo-ET reconstruction of an intact human NPC. The symmetric core, embedded in the nuclear envelope, is decorated with NUP358 (red) domains bound to Ran (gray), flexibly projected into the cytoplasm, and CFNCs (pink) overlooking the central transport channel.

CONCLUSION: We have substantially advanced the biochemical and structural characterization of the asymmetric nucleoporins' architecture and attachment at the cytoplasmic and nuclear faces of the NPC. Our near-atomic composite structure of the human NPC's cytoplasmic face provides a biochemical and structural framework for elucidating the molecular basis of mRNP remodeling, viral virulence factor interference with NPC function, and the underlying mechanisms of nucleoporin diseases at the cytoplasmic face of the NPC. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: hoelz@caltech.edu
†These authors contributed equally to this work.
Cite this article as C. J. Bley *et al.*, *Science* **376**, eabm9129 (2022). DOI: 10.1126/science.abm9129

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abm9129

RESEARCH ARTICLE

NUCLEAR PORE COMPLEX

Architecture of the cytoplasmic face of the nuclear pore

Christopher J. Bley[1]†, Si Nie[1]†, George W. Mobbs[1]†, Stefan Petrovic[1]†, Anna T. Gres[1]†‡, Xiaoyu Liu[1]†§, Somnath Mukherjee[2], Sho Harvey[1], Ferdinand M. Huber[1]¶, Daniel H. Lin[1]#, Bonnie Brown[1], Aaron W. Tang[1], Emily J. Rundlet[1]††‡‡, Ana R. Correia[1]§§, Shane Chen[3], Saroj G. Regmi[3], Taylor A. Stevens[1], Claudia A. Jette[1], Mary Dasso[3], Alina Patke[4], Alexander F. Palazzo[5], Anthony A. Kossiakoff[2], André Hoelz[1]*

The nuclear pore complex (NPC) is the sole bidirectional gateway for nucleocytoplasmic transport. Despite recent progress in elucidating the NPC symmetric core architecture, the asymmetrically decorated cytoplasmic face, essential for messenger RNA (mRNA) export and a hotspot for nucleoporin-associated diseases, has remained elusive. Here we report a composite structure of the human cytoplasmic face obtained by combining biochemical reconstitution, crystal structure determination, docking into cryo–electron tomographic reconstructions, and physiological validation. Whereas species-specific motifs anchor an evolutionarily conserved ~540-kilodalton heterohexameric cytoplasmic filament nucleoporin complex above the central transport channel, attachment of the NUP358 pentameric bundles depends on the double-ring arrangement of the coat nucleoporin complex. Our composite structure and its predictive power provide a rich foundation for elucidating the molecular basis of mRNA export and nucleoporin diseases.

The sequestration of genetic material in the nucleus represents one of the hallmarks of evolution but creates the necessity for selective bidirectional transport across the nuclear envelope (*1*–*4*). The nuclear pore complex (NPC) is the sole gateway through which folded proteins and protein–nucleic acid complexes cross the nuclear envelope, making this transport organelle an essential machine for all eukaryotic life. Besides its direct role as a transport channel, the NPC serves as an organizer for nuclear and cytoplasmic processes that are essential for the flow of genetic information from DNA to RNA to protein, including transcription, spliceosome assembly, mRNA export, and ribosome assembly (*1*–*4*). Dysfunction of the NPC or its components represents a major cause of human disease (*2*, *5*, *6*).

Architecturally, the NPC consists of a central core with eightfold rotational symmetry across a nucleocytoplasmic axis and twofold rotational symmetry across the plane of the nuclear envelope, which links to compartment-specific asymmetric cytoplasmic filaments (CFs) and a nuclear basket structure (Fig. 1A) (*1*, *2*). The NPC is built from ~34 different proteins, termed nucleoporins (nups), that are organized into distinct subcomplexes. Multiple copies of each nup in the NPC add up to an assembly that reaches a molecular mass of ~110 MDa in vertebrates. The symmetric core of the NPC is composed of an inner ring and two spatially segregated outer rings. The inner ring is embedded in nuclear envelope pores generated by the circumscribed fusion of the double membrane of the nuclear envelope. The diffusion barrier is formed by unstructured phenylalanine-glycine (FG) repeats that fill the central transport channel, imposing a gradually increasing barrier to passive diffusion of macromolecules >40 kDa (*1*–*4*). Transport factors, collectively termed karyopherins, overcome the diffusion barrier by binding to FG repeats, thereby transporting cargo across the nuclear envelope (*7*–*9*). A substantial fraction of the FG repeats in the inner ring is contributed by a heterotrimeric channel nup complex (CNT), which is anchored by a single assembly sensor motif (*10*–*12*). The outer rings sit atop the nuclear envelope, sandwiching the inner ring from both sides. The outer rings are primarily formed by the Y-shaped coat nup complex (CNC; also referred to as the Y-complex or the Nup107-160 complex) and serve as a platform for asymmetric incorporation of the CF and nuclear basket nups.

Two decades ago, the atomic-level characterization of the NPC began with individual nup domains and progressed to nup complexes of increasing size and complexity, culminating in the ~400-kDa heteroheptameric CNC (*11*, *13*–*28*). Simultaneously, advances in cryo–electron tomography (cryo-ET) data acquisition and processing gradually increased the resolution of intact NPC three-dimensional (3D) reconstructions (*29*). Docking of the CNC into a ~32-Å cryo-ET map of the intact human NPC demonstrated that two reticulated eight-membered CNC rings, linked by head-to-tail interactions, are present on each side of the nuclear envelope (*27*, *30*). Moreover, this advance established that the resolution gap between high- and low-resolution structural methods can be overcome by combining biochemical reconstitution and x-ray crystallographic characterization of nups with cryo-ET reconstruction of the intact NPC. Expansion of this approach to the nine nups constituting the inner ring rapidly led to the reconstitution of two distinct ~425-kDa inner-ring complexes (IRCs) and the elucidation of their components' structures (*10*–*12*, *20*, *31*–*38*). In turn, this advance enabled the determination of the near-atomic composite structure of the entire ~56-MDa symmetric core of the human NPC, establishing the stoichiometry and placement of all 17 symmetric nups within a ~23-Å cryo-ET reconstruction (*38*, *39*). Subsequently, the architecture of the *Saccharomyces cerevisiae* NPC was determined by means of a similar approach, using high-resolution nup crystal structures and ~25-Å cryo-ET maps of either detergent-purified or in situ NPCs (*40*, *41*). Compared with the human NPC, the *S. cerevisiae* NPC lacks the distal CNC ring and associated nups on both sides of the nuclear envelope, but the relative nup arrangement within the rest of the symmetric core remains essentially identical (*38*, *39*, *42*).

Projecting from the cytoplasmic face of the NPC, the CF nups recruit cargo•transport factor complexes for nucleocytoplasmic transport and orchestrate the export and remodeling of messenger ribonucleoprotein particles (mRNPs) in preparation for translation (*2*, *43*). The nine-component CF nup machinery represents a hotspot for human diseases ranging from degenerative brain disorders and cardiac diseases to cancer (*2*, *5*, *6*). Although linked to the human CF nups NUP358, NUP214, NUP62, NUP88, NUP98, GLE1, NUP42, RAE1, and DDX19, the pathophysiology and optimal

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA. [2]Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637, USA. [3]Division of Molecular and Cellular Biology, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA. [4]Division of Biology and Biological Engineering, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA. [5]Department of Biochemistry, University of Toronto, Toronto, ON M5G 1M1, Canada.
***Corresponding author. Email: hoelz@caltech.edu**
†These authors contributed equally to this work. ‡Present address: Clinical Research Methodology, Scientific Solutions, Worldwide Clinical Trials, 600 Park Offices Drive Suite 200, Research Triangle Park, NC 27709, USA. §Present address: Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, 609 Charles E. Young Drive East, Los Angeles, CA 90095, USA. ¶Present address: Odyssey Therapeutics, Inc., Industriepark Höchst G875, 65926 Frankfurt am Main, Germany. #Present address: Whitehead Institute for Biomedical Research, 455 Main Street Cambridge, MA 02142, USA. ††Present address: Department of Structural Biology, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA. ‡‡Tri-Institutional PhD Program in Chemical Biology, Weill Cornell Medicine, 1300 York Avenue, New York, NY 10065, USA. §§Present address: Amgen Research, Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320, USA.

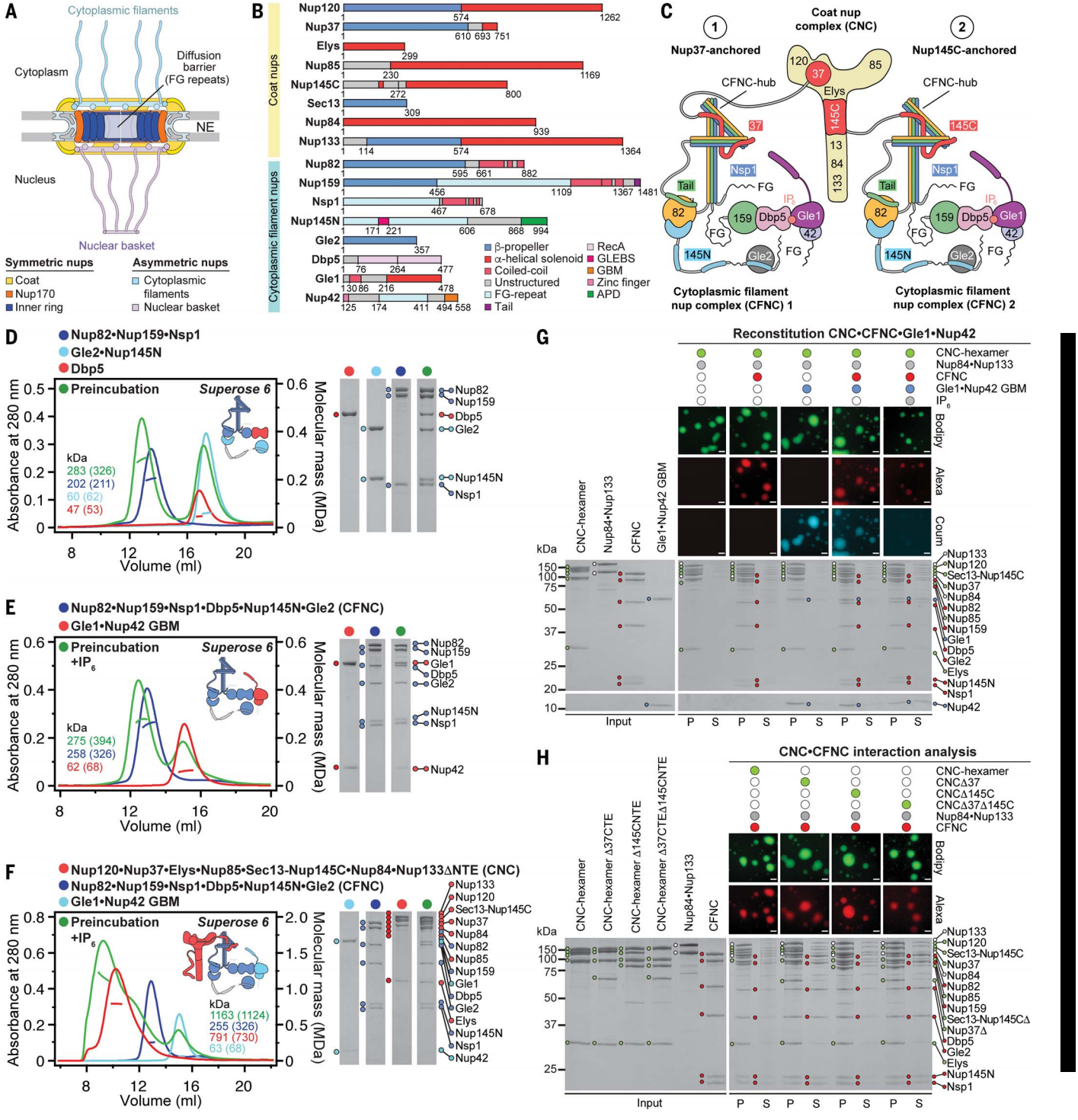


Fig. 1. Reconstitution of a 16-protein *C. thermophilum* complex composed of coat and CF nups. (**A**) Cross-sectional schematic of the fungal NPC architecture. NE, nuclear envelope. (**B**) Domain structures of the coat and CF nups. (**C**) Schematic representation summarizing our biochemical reconstitution and dissection experiments with purified recombinant *C. thermophilum* nups, illustrating the CFNC architecture and its attachment to the CNC. The CNC harbors two assembly sensors, Nup37^{CTE} and Nup145C^{NTE}, each anchoring a CFNC via its central hub, with Nup37^{CTE} exhibiting tighter binding than Nup145C^{NTE}. (**D** to **F**) SEC-MALS interaction analyses, showing the stepwise biochemical reconstitution starting with (D) the CFNC (green) from Nup82•Nup159•Nsp1 (blue), Gle2•Nup145N (cyan), and Dbp5 (red); then (E) CFNC•Gle1•Nup42^{GBM} (green) from CFNC (blue) and Gle1•Nup42^{GBM} (red); and culminating with (F) the 16-protein CNC•CFNC•Gle1•Nup42^{GBM} complex (green) from CNC (red), CFNC (blue), and Gle1•Nup42^{GBM} (cyan). SDS-PAGE gel strips of peak fractions are shown. Measured molecular masses are indicated, with respective theoretical masses in parentheses. (**G** and **H**) LLPS interaction assays, assessing (G) CFNC (red) and Gle1•Nup42^{GBM} (cyan) incorporation into CNC-LLPS (green), and (H) CFNC incorporation into CNC-LLPS, lacking either one or both Nup37^{CTE} and Nup145C^{NTE} assembly sensors. N-terminally fluorescently labeled CNC (Bodipy), CFNC (Alexa Fluor 647), and Gle1•Nup42^{GBM} (Coumarin) were visualized by fluorescence microscopy. Pelleted CNC condensate phase (P) and soluble (S) fractions were analyzed by SDS-PAGE and visualized by Coomassie brilliant blue staining. Scale bars, 10 μm.

therapeutic strategies for these conditions remain ill defined.

Here we present insight into the atomic and higher-order architecture, function, and mechanism of action of the CF nups in the human and thermophilic fungus *Chaetomium thermophilum* NPCs. First, we uncover a conserved modular architecture within the heterohexameric CF nup complex (CFNC) of both species: Holding the CFNC together is a coiled-coil hub built like the CNT but formed by NUP62 with the C-terminal regions of NUP88 and NUP214, while their N-terminal β-propeller domains link to the mRNA export factors NUP98•RAE1 and the DEAD-box RNA helicase DDX19, respectively, which in turn recruit the remaining complex components. We further uncover evolutionarily divergent mechanisms for the attachment of the intact CFNC at the cytoplasmic face of the NPC, which in *C. thermophilum* involves two distinct assembly sensors in the CNC that do not exist in humans. We assemble the *C. thermophilum* CNC and CF nups into a ~1.1-MDa 16-protein complex and find that it can be remodeled by inositol hexaphosphate (IP_6). Toward dissecting the molecular mechanism of mRNA export, we systematically characterize the propensity of CF nups for RNA binding and find previously unidentified capabilities in two CFNC subcomplexes (GLE1•NUP42 and NUP88•NUP214•NUP98) as well as different parts of the metazoan-specific NUP358. To build a composite structure of the human NPC cytoplasmic face, we determine crystal structures of the $NUP88^{NTD}$•$NUP98^{APD}$ complex and all remaining structurally uncharacterized regions of NUP358, uncovering a previously unobserved S-shaped fold of three α-helical solenoids of the NUP358 N-terminal domain as well as a complex mechanism for NUP358 oligomerization. Docking of the newly identified structures, along with previously characterized CF nups, into a previously reported ~23-Å cryo-ET map and a new ~12-Å cryo-ET map of the intact human NPC (provided by the Beck group), as well as an ~8-Å region of an anisotropic single-particle cryo–electron microscopy (cryo-EM) composite map of the *Xenopus laevis* cytoplasmic NPC face, accounts for all of the asymmetric density on the cytoplasmic NPC side resolved in the maps (*44*–*46*). Validating our quantitative docking analysis in human cells engineered to enable rapid, inducible NUP358 depletion, we surprisingly find NUP358 to be dispensable for the architectural integrity of the assembled interphase NPC and mRNA export but to have a general role in translation. Docking of the CFNC hub in close proximity to a NUP93 fragment that, in the inner ring, acts as the assembly sensor for the CNT allows us to predict and experimentally confirm that NUP93 also recruits the structurally related CFNC on the cytoplasmic face, thereby enabling identification of the elusive human CFNC NPC anchor. Thus, our near-atomic composite structure has predictive power, demonstrating its general utility for the mechanistic dissection of essential cellular events that occur on the cytoplasmic face of the human NPC.

Results

Modular architecture of the evolutionarily conserved six-protein CFNC

Although pairwise interactions between selected CF nups had previously been reported, comprehensive knowledge on the entire CF nup interaction network has remained unavailable (*11*, *47*–*64*). Using nups from the thermophilic fungus *C. thermophilum*, which exhibit superior biochemical stability, we previously elucidated the interaction network of the 17 symmetric core nups (*38*). Therefore, we first sought to establish the protein-protein interaction network and complex stoichiometry of the eight evolutionarily conserved *C. thermophilum* CF nups: Nup159, Nup82, Nsp1, Nup145N, Gle2, Dbp5, Gle1, and Nup42 (Fig. 1B and fig. S1) (*2*). Most CF nups contain both structured and unstructured regions that can harbor multiple distinct binding sites and FG repeats. We established expression and purification protocols for the *C. thermophilum* CF nups, omitting FG-repeat regions as well as an unstructured linker region in Nup145N to improve solubility, and analyzed their binding by size-exclusion chromatography coupled with multiangle light scattering (SEC-MALS) and a liquid-liquid phase separation (LLPS) interaction assay (Fig. 1, figs. S1 to S26, and tables S1 to S6). For a detailed description of these experiments, see the supplementary text.

Mixture of Nup82•Nup159•Nsp1 with Gle2•Nup145N and Dbp5 results in the formation of a stoichiometric heterohexameric CFNC (Fig. 1, C and D, and fig. S4A), which is held together by a parallel coiled-coil heterotrimer formed by the C-terminal domains of Nup82•Nup159•Nsp1, termed the CFNC hub (figs. S4 to S6). The CFNC is tethered to the NPC by two mutually exclusive assembly sensors targeting the CFNC hub. These anchor points are located within primarily unstructured regions [N-terminal extensions (NTEs) and C-terminal extensions (CTEs)] present in the CNC constituents $Nup37^{CTE}$ and $Nup145C^{NTE}$, which supply a strong and weak binding site, respectively, permitting two CFNCs to bind a single CNC (figs. S7 to S17). The Gle1•Nup42 complex has also been shown to locate at the cytoplasmic face of the NPC, forming an IP_6-dependent interaction with Dbp5 (*53*, *58*, *63*, *65*). We demonstrate stoichiometric incorporation of Gle1•Nup42 into both the CFNC and CNC•CFNC complexes in the presence of IP_6 (Fig. 1, E and F, and figs. S18 to 23). Additionally, we identify an interaction formed between Gle1•Nup42 and the CNC that is disrupted upon addition of IP_6, establishing that the CNC-CF nup interaction network can be remodeled (figs. S23 to S25).

Given the special importance of the CF nups in human disease, we next tested whether the molecular architecture of the CFNC is evolutionarily conserved from *C. thermophilum* to humans. The human CFNC is composed of NUP88, NUP214, NUP62, NUP98, RAE1, and DDX19. Apart from a rearrangement of the FG-repeat and coiled-coil regions in NUP214, the domain organization of the human CFNC nups is identical to that of the *C. thermophilum* orthologs (Fig. 2 and fig. S27). Indeed, mixing the NUP88•NUP214•NUP62 heterotrimer with RAE1•NUP98 and DDX19 resulted in a stoichiometric *Homo sapiens* CFNC heterohexamer (Fig. 2C and figs. S28 and S29). Similarly, a systematic pairwise interaction analysis established that the modular CFNC architecture characterized in *C. thermophilum* is conserved in humans (Fig. 2, D and E, and figs. S30 to S39).

Together, our data establish that the CF nups form an evolutionarily conserved six-protein complex held together by an extensive parallel coiled-coil hub generated by the C-terminal regions of Nup82/NUP88, Nup159/NUP214 and Nsp1/NUP62, which shares architectural similarities with the heterotrimeric Nsp1/NUP62•Nup49/NUP58•Nup57/NUP54 CNT (*11*). The Nup82/NUP88 N-terminal β-propeller domain is attached by an interaction between the C-terminal α-helical TAIL fragment of Nup159/NUP214 and provides a binding site for the Nup145N/NUP98 autoproteolytic domain (APD), which in turn recruits Gle2/RAE1 to the NPC. Analogously, the Nup159/NUP214 N-terminal β-propeller domain provides a binding site for the DEAD-box helicase Dbp5/DDX19. In *C. thermophilum*, the CFNC hub is anchored to the CNC by two distinct assembly sensors in $Nup37^{CTE}$ and $Nup145C^{NTE}$, similar to the anchoring of the CNT by the $Nic96^{R1}$/$NUP93^{R1}$ assembly sensor in the inner ring. By contrast, the human CNC lacks comparable assembly sensor motifs, which suggests that there are alternative mechanisms for anchoring CF nups at the cytoplasmic face of the human NPC.

RNA interactions of human CF nups

Given their essential roles in mRNA export, we next sought to identify which of the human CF nups had RNA-binding capabilities (*50*, *66*–*69*). Previous work that used disparate methods and diverse but inconsistent probes had established DDX19 and $RAE1•NUP98^{GLEBS}$ binding to U_{10} single-stranded RNA (ssRNA), degenerate ssRNA, poly(A), poly(C), poly(G) RNA, as well as ssDNA and double-stranded DNA (dsDNA) across a variety of assays (*50*, *56*, *57*, *63*, *70*).

Taking advantage of our complete set of purified human CF nup domains and subcomplexes,

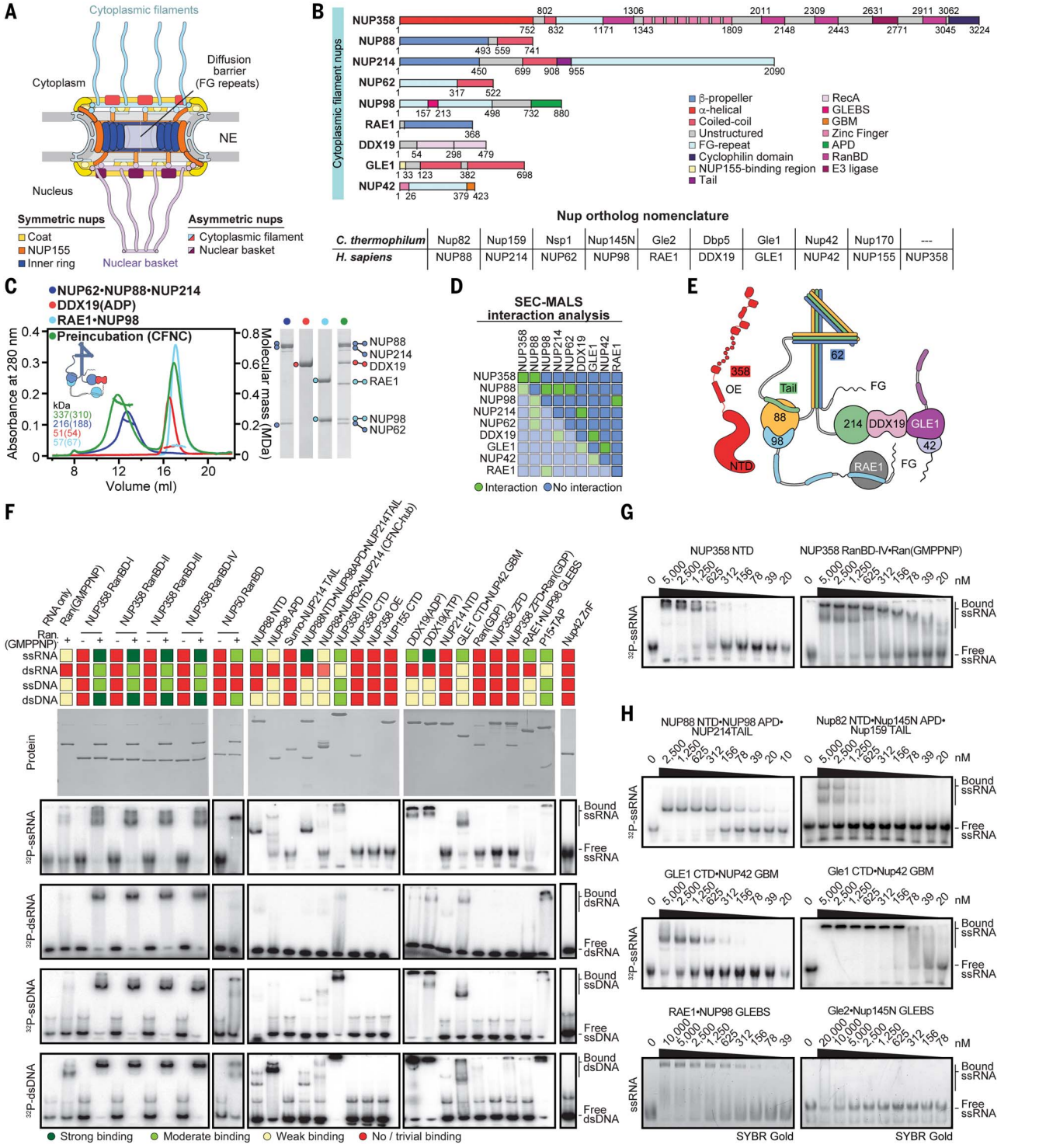


Fig. 2. Conserved modular architecture and RNA-binding properties of the human CF nups. (**A**) Cross-sectional schematic of the human NPC architecture. (**B**) Domain structures of human CF nups. Nomenclature of *H. sapiens* and *C. thermophilum* nup orthologs is indicated. (**C**) Biochemical reconstitution of the ~310-kDa heterohexameric human CFNC. SEC-MALS interaction analysis of NUP88•NUP214•NUP62 (blue), DDX19 (ADP) (red), RAE1•NUP98 (cyan), and their preincubation (green). Measured molecular masses are indicated, with theoretical masses in parentheses. SDS-PAGE gel strips of peak fractions visualized by Coomassie brilliant blue staining are shown. (**D**) Summary of pairwise SEC-MALS interaction analyses between human CF nups. (**E**) Schematic summary of the human CFNC architecture and the CF nup interaction network. (**F**) Human CF nup domains and complexes were assayed for binding to ssRNA and dsDNA probes by EMSA. Input proteins resolved by SDS-PAGE were visualized by Coomassie brilliant blue staining. Qualitative assessment of nucleic acid binding is denoted by color-coded boxes. (**G** and **H**) EMSAs with ssRNA titrated against (G) metazoan-specific NUP358^{NTD} and NUP358$^{RanBD\text{-}IV}$•Ran(GMPPNP), and (H) the indicated *H. sapiens* CFNC subcomplexes and their *C. thermophilum* orthologs.

we carried out a comprehensive electrophoretic mobility shift assay (EMSA) screen to systematically assess binding against a consistent set of ssRNA and dsRNA probes (Fig. 2F). In addition to the established RNA binders DDX19 and RAE1•NUP98, we identified ssRNA binding by GLE1^{CTD}•NUP42^{GBM} and NUP88^{NTD}, an activity enhanced in the context of the NUP88^{NTD}•NUP98^{APD}•NUP214^{TAIL} complex. We tested their *C. thermophilum* orthologs and found these RNA binding activities to be conserved (Fig. 2, G and H, and fig. S40). Next, we analyzed the metazoan-specific NUP358. We detected moderate RNA binding for the NUP358 N-terminal domain (NTD) (Fig. 2F and fig. S40G). Additionally, we found that the four NUP358^{RanBD}•Ran(GMPPNP) complexes preferentially bound ssRNA (Fig. 2F and fig. S40E). Additional details of RNA binding can be found in the supplementary text. Future studies are needed to delineate whether these RNA-binding sites present sequence-specific RNA affinity and what the implications of such specificity would be in the overall mRNA export pathway.

Structural and biochemical analyses of NUP358

NUP358 is a 3224-residue metazoan-specific CF component and the largest constituent of the NPC (*71*–*74*). Previous studies established that its N-terminal ~900-residue α-helical region is necessary for nuclear envelope recruitment (*75*). Within this region, the first 145 residues have been biochemically and structurally characterized, shown to form three tetratricopeptide repeats (TPRs) (*76*). Guided by secondary structure predictions, we systematically screened expression constructs for solubility, identifying three fragments: NUP358$^{NTD\Delta TPR}$ (residues 145 to 752), NUP358^{NTD} (residues 1 to 752), and NUP358$^{1\text{-}832}$ (an extended region spanning residues 1 to 832). Subsequent purifications revealed that the NUP358^{NTD} and NUP358$^{1\text{-}832}$ fragments behave differently, with the latter forming amorphous precipitates in buffers with NaCl concentrations below 300 mM. Therefore, we characterized these NUP358 fragments in both high-salt (350 mM NaCl) and low-salt (100 mM NaCl) buffers wherever possible.

NUP358^{NTD} exhibited concentration-dependent homodimerization in low-salt buffer, with measured molecular masses between values corresponding to monomeric and dimeric species, but existed as a monomeric species in high-salt buffer (Fig. 3, A and B, and fig. S41). Conversely, NUP358$^{NTD\Delta TPR}$ was exclusively monomeric, suggesting that the TPR mediates homodimerization (Fig. 3B and fig. S41). Furthermore, the extended NUP358$^{1\text{-}832}$ fragment forms oligomers with measured molecular masses between those of a tetramer and a pentamer (Fig. 3C and fig. S41). Subsequent C-terminal mapping revealed an oligomerization element (OE) between residues 802 and 832, forming salt-insensitive concentration-dependent oligomers between dimers and tetramers (Fig. 3G and fig. S42). Thus, NUP358 oligomerization is mediated by the TPR and OE regions, located on opposite sides of the N-terminal α-helical region.

To aid the crystallization of the entire NUP358^{NTD}, we generated high-affinity synthetic antibody fragments (sABs) by phage display selection (*77*). By systematically screening the generated 62 sABs as crystallization chaperones, we identified a NUP358^{NTD}•sAB-14 complex that crystallized, enabling de novo structure determination of the entire NUP358^{NTD} at 3.95-Å resolution (tables S7 to S10). To unambiguously assign the Nup358^{NTD} sequence register, we crystallized 17 seleno-L-methionine mutants (fig. S43 and tables S11 and S12).

The asymmetric unit contained two copies of the NUP358^{NTD}•sAB-14 complex, in one of which the first three and a half TPR repeats are not resolved. The second copy forms extensive interactions with a symmetry-related molecule (Fig. 3, D to F, and fig. S44). This NUP358^{NTD} dimer reveals two alternative TPR conformations in which the TPR either forms a continuous N-terminal solenoid (open) or folds back, separating TPR4 and forming electrostatic interactions with HEAT repeats 5 to 7 of the N-terminal solenoid (closed) (Fig. 3F, fig. S44C, and Movie 1). Toggling between these two states provides a molecular explanation for the salt-sensitive, concentration-dependent dimerization behavior of NUP358^{NTD} (Fig. 3B). Because the open confirmation is the one identified in the intact NPC (see below), we focus our description on this state.

The open conformation of NUP358^{NTD} can be divided in three sections: an N-terminal α-helical solenoid composed of four TPRs and four HEAT repeats, a central α-helical wedge domain, and a short C-terminal α-helical solenoid formed by three HEAT repeats (fig. S44D). The N- and C-terminal TPR and HEAT repeats are capped by solvating helices. Inserted between α17 of the N-terminal solenoid and α20 of the wedge domain is a ~50-residue loop that wraps around the convex face of the N-terminal solenoid. The N-terminal solenoid and wedge domain form a composite concave surface with a pronounced overall positive charge (figs. S45 and S46). The central wedge domain makes extensive hydrophobic contacts with the sides of the N- and C-terminal solenoids, generating a noncanonical S-shaped architecture (fig. S44D). Indeed, a Dali 3D search of the Protein Data Bank (PDB) revealed that the NUP358^{NTD} architecture has not been observed previously (*78*).

Our biochemical analysis revealed that NUP358^{NTD} interacts weakly with NUP88^{NTD} and has RNA-binding activity, both of which were salt sensitive (figs. S47 and S48). By splitting NUP358^{NTD} into two fragments, NUP358^{TPR} and NUP358$^{NTD\Delta TPR}$, we show that both halves are necessary yet insufficient for either NUP88^{NTD} or RNA binding (figs. S47 and S48). To further map these interactions, we performed a saturating NUP358^{NTD} surface mutagenesis, screening 106 mutants for NUP88^{NTD} and RNA binding (fig. S49). We found that positively charged residues in the concave surface mediate binding to both NUP88^{NTD} and RNA. Mutations that abolished NUP88^{NTD} binding clustered exclusively on the N-terminal solenoid, whereas RNA disruption required additional mutations in the wedge domain. By systematically combining individual alanine substitutions, we identified a NUP358^{NTD} 2R5K mutation, which abolished both interactions (fig. S50).

Next, we determined the crystal structure of NUP358^{OE} at 1.1-Å resolution (table S13). NUP358^{OE} is a small α-helical element that homotetramerizes to form an antiparallel bundle (Fig. 3, G and H, and fig. S42A). The core of the α-helical bundle is lined with hydrophobic residues that coordinate oligomeric interhelical packing, demonstrated by the monomeric form assumed by the NUP358^{OE} LIQIML mutant (Fig. 3G; fig. S42, B to E; and Movie 2). To validate our NUP358^{OE} structure, we tested the effect of introducing the LIQIML mutation into the larger NUP358$^{NTD\text{-}OE}$. Whereas wild-type NUP358$^{NTD\text{-}OE}$ formed higher-order oligomeric species, the oligomerization profile of the LIQIML NUP358$^{NTD\text{-}OE}$ mutant matched that of the OE-less NUP358^{NTD}, presenting concentration-dependent dimerization in low-salt buffer but persisting in a monodisperse monomeric state in high-salt buffer (fig. S42, F and G).

Our data show that NUP358^{NTD} is composed of three distinct α-helical solenoids that interact in a previously unobserved manner, adopting a distinctive overall S-shaped architecture with a propensity to form domain-swapped homodimers. Connected to NUP358^{NTD} by a ~50-residue linker is an oligomerization element that forms homotetramers/pentamers in solution. These dual modes of homo-oligomerization provide a plausible explanation for NUP358's propensity to form phase separation, as observed during NPC assembly in *Drosophila melanogaster* oocytes (*79*).

Ran interactions with human asymmetric nups

Nucleocytoplasmic transport depends on karyopherin transport factors (Kaps) with directionality imposed by a cellular gradient of the small guanosine triphosphatase (GTPase) Ran; nuclear Ran(GTP) is elevated by a factor of ~200 compared with the primarily cytoplasmic Ran(GDP) (*2*, *7*, *9*). Multiple Ran-binding sites are distributed among the asymmetric nups at the cytoplasmic and nuclear

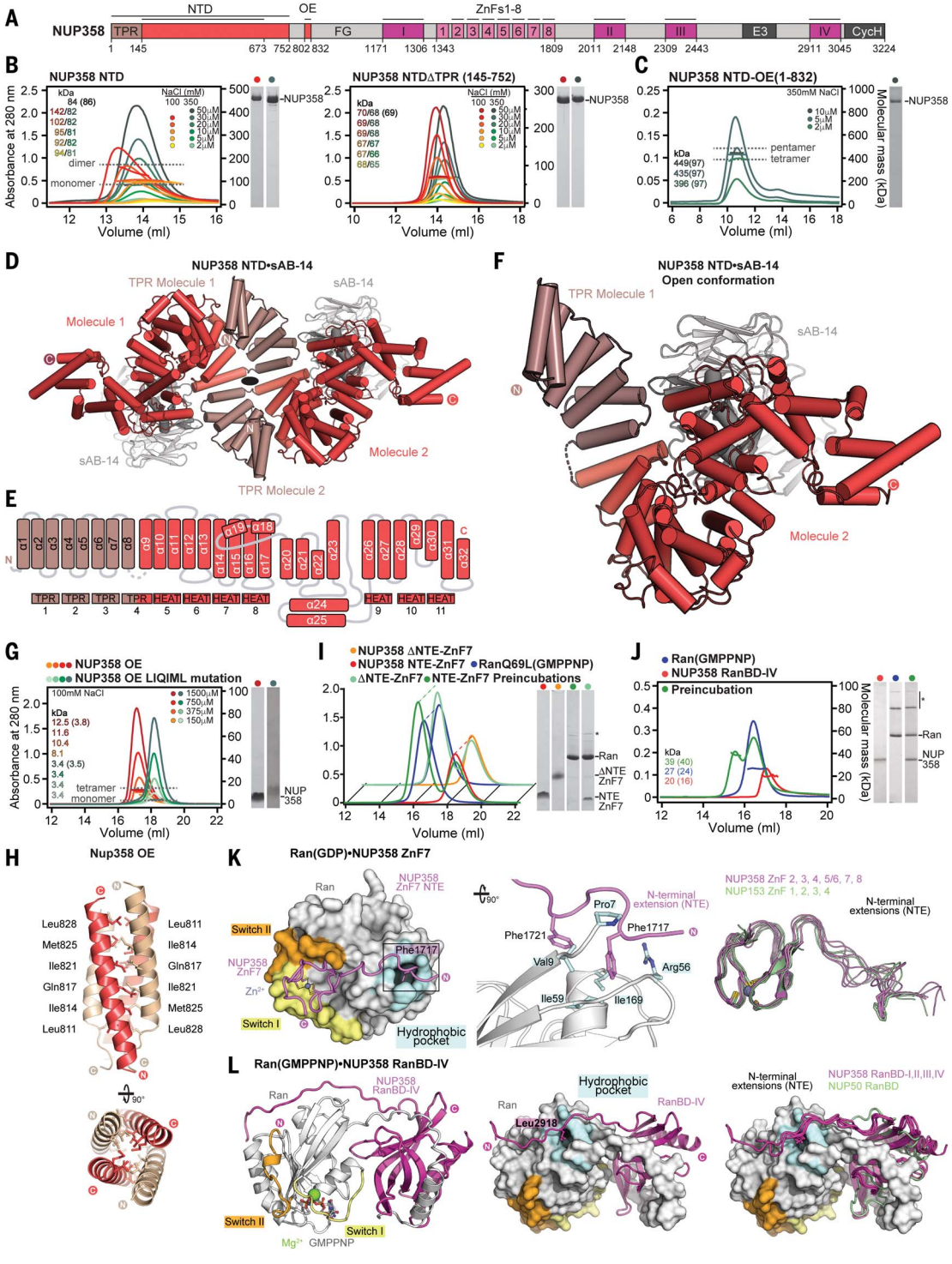
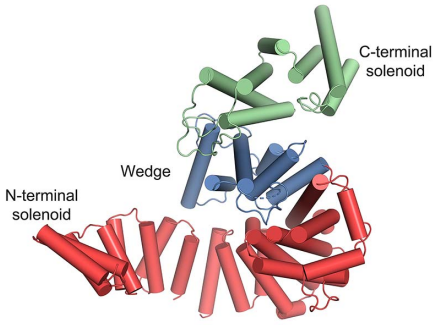


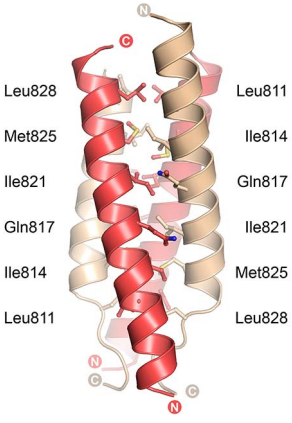
Fig. 3. Structural analysis and biochemical characterization of NUP358. (**A**) Domain structure of NUP358. Black lines indicate boundaries of the crystallized fragments. (**B** and **C**) SEC-MALS analysis of the oligomeric behavior of (B) $NUP358^{NTD}$ and $NUP358^{NTD\Delta TPR}$ and (C) $NUP358^{NTD\text{-}OE}$, performed at the indicated protein concentrations. Measured molecular masses are indicated, with theoretical masses in parentheses. SDS-PAGE gel strips of peak fractions are shown and visualized by Coomassie brilliant blue staining. (**D**) Cartoon representation of the $NUP358^{NTD}$•sAB-14 cocrystal structure dyad of the $P6_522$ lattice, illustrating the $NUP358^{NTD}$ dimer between symmetry-related molecules. (**E**) Schematic of the $NUP358^{NTD}$ structure and structural motifs. (**F**) TPR of molecule 1 complements α-helical stacking of the N-terminal solenoid of the symmetry-related molecule 2, generating the open conformation of $NUP358^{NTD}$. (**G**) SEC-MALS analysis of the oligomeric behavior of $NUP358^{OE}$ and the $NUP358^{OE}$ LIQIML mutant performed at the indicated protein concentrations. (**H**) Cartoon representation of the homotetrameric $NUP358^{OE}$ crystal structure with hydrophobic core residues shown in ball-and-stick representation. (**I**) SEC interaction analysis of $NUP358^{ZnF7}$ and $NUP358^{ZnF7\Delta NTE}$ binding to Ran(GTP). (**J**) SEC-MALS interaction analysis of $NUP358^{RanBD\text{-}IV}$ binding to Ran(GMPPNP). (**K**) Cocrystal structure of $NUP358^{ZnF7}$•Ran(GDP), shown in cartoon and surface representation (left). The inset indicates the location of the magnified and 90°-rotated view of the Ran hydrophobic pocket (middle). Superposition of the six $NUP358^{ZnF}$•Ran(GDP) and four $NUP153^{ZnF}$•Ran(GDP) cocrystal structures with the Zn^{2+}-coordinating cysteines and Ran-burying NTE hydrophobic residues shown as sticks (right). (**L**) Cocrystal structure of $NUP358^{RanBD\text{-}IV}$•Ran(GMPPNP) with $NUP358^{RanBD\text{-}IV}$ shown in cartoon representation and Ran(GMPPNP) shown in cartoon (left) or surface (middle) representation. Superposition of Ran(GMPPNP) bound to $NUP358^{RanBD\text{-}I}$, $NUP358^{RanBD\text{-}II}$, $NUP358^{RanBD\text{-}III}$, $NUP358^{RanBD\text{-}IV}$, and $NUP50^{RanBD}$ (right).

sides of the NPC in the form of distinct Ran-binding domains (RanBDs) and Zn^{2+}-finger (ZnF) modules. On the cytoplasmic face, NUP358 contains four dispersed RanBDs and a central zinc finger domain (ZFD) with a tandem array of eight ZnFs (Fig. 3A) (*80*). On the nuclear side, NUP153 and NUP50 contain a central ZFD with four ZnFs and a solitary C-terminal RanBD, respectively (fig. S51) (*81*, *82*).

By testing the Ran(GDP/GTP)-binding activity of all 17 domains by SEC-MALS, we confirmed that all domains bound to Ran, as expected, except for $NUP358^{ZnF1}$ (Fig. 3, I and J, and figs. S51 to S55). Consistent with



Movie 1. Structure of NUP358^{NTD}. A 360° rotation of the NUP358^{NTD}•sAB-14 cocrystal structure, illustrating the NUP358^{NTD} dimer between symmetry-related molecules, followed by a comparison of the two possible TPR conformations giving rise to the open and closed states, concluding with a 360° rotation of the NUP358^{NTD} open confirmation.



Movie 2. Structure of NUP358^{OE}. A 360° rotation of the of the homotetrameric NUP358^{OE} crystal structure, with hydrophobic core residues shown in ball-and-stick representation followed by an end-on view.

previous reports, the RanBDs of NUP358 and NUP50 only bound Ran(GTP), whereas the ZnFs in NUP358 and NUP153 bound Ran in both nucleotide states but showed a preference for Ran(GDP) (figs. S52 and S54) (*80*, *83*–*85*). To clarify the molecular basis for the differential binding behaviors, we determined the cocrystal structures of all 16 domains bound to Ran in their preferred nucleotide-bound state at 1.8- to 2.45-Å resolution (figs. S51 and S55, Movies 3 and 4, and tables S14 to S16). For an expanded description of these structures, see the supplementary text.

Together, our data establish that the human CF and nuclear basket nups NUP358, NUP153, and NUP50 harbor a total of 16 distinct Ran-binding sites that, given their stoichiometry in the NPC, could together recruit up to several hundred Ran molecules. Considering the substantial size difference between metazoan and *S. cerevisiae* cells, it is conceivable that additional Ran-binding sites provided by the metazoan-specific asymmetric nups NUP358 and NUP153 help ensure that Ran concentrations in the NPC vicinity are high enough to enable nucleocytoplasmic transport, as has been previously suggested (*85*).

Docking of NUP358^{NTD} into the cytoplasmic unassigned density cluster I

NUP358 is known to reside on the cytoplasmic face of the NPC. Without a structure, its location in the NPC could previously be inferred only from the differential absence of unassigned density in an ~38-Å cryo-ET map of a NUP358-depleted human NPC (*44*). In the accompanying manuscript, we describe how quantitative docking of residue-level resolution structures into the symmetric core of a ~12-Å cryo-ET map of the intact human NPC led to assignment of 16 copies of the symmetric nups NUP205 and NUP93 in the cytoplasmic outer ring, as well as the identification of two clusters (I and II) of unassigned density (Fig. 4A), of which the first corresponds to the previously observed NUP358-dependent density (*42*). Because of the large size and distinctive fold of our newly elucidated NUP358^{NTD} crystal structure, we sought to directly determine its position in the intact human NPC (Fig. 4A).

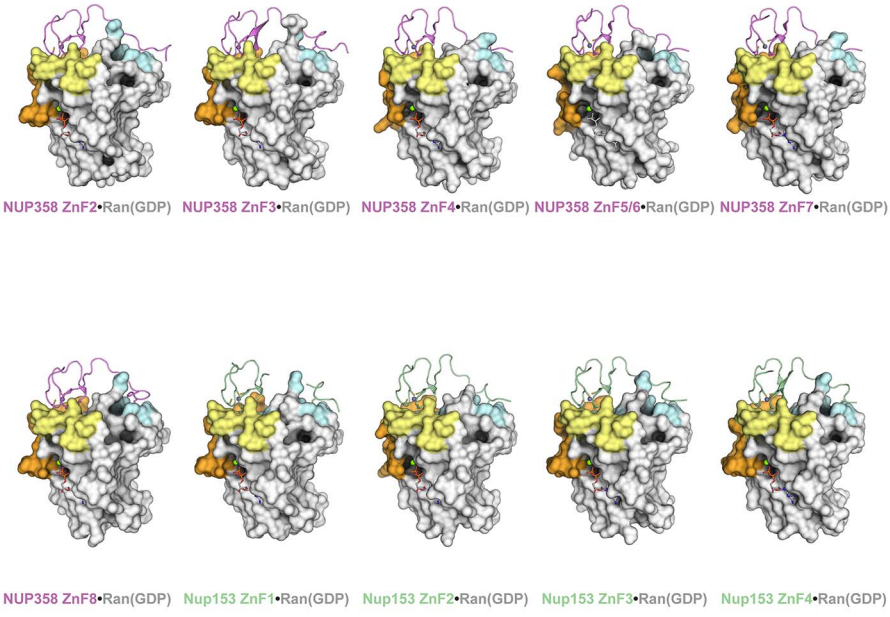
In our docking analysis, we calculated correlations between a new ~12-Å cryo-ET map of the intact human NPC (provided by the Beck group) and 1 million resolution-matched densities simulated from either the open or closed conformation of the NUP358^{NTD} crystal structure, randomly placed and locally fit-optimized in the asymmetric unit of the full ~12-Å cryo-ET map (*46*). Unlike for the closed NUP358^{NTD} conformation, docking scores for five placements of the open NUP358^{NTD} conformation segregated to high confidence of placement and located to the previously unassigned density cluster I, leaving no unexplained density (fig. S56). We found four copies of NUP358^{NTD} to be interfaced with the α-helical solenoid folds of the CNC components NUP96, NUP107, and the distal copy of NUP93^{SOL}, wrapping around the stalks of the tandem-arranged Y-shaped CNCs in pairs, at equivalent distal and proximal positions (Fig. 4, B and C). As identified in the docking analysis of the symmetric core reported in the accompanying manuscript, the distal NUP93^{SOL} bisects the stalks of the tandem-arranged Y-shaped CNCs by interfacing with the distal NUP107 and the proximal NUP96 α-helical solenoids, cloistered between the four NUP358^{NTD} copies (Fig. 4D) (*42*). Lastly, the fifth NUP358^{NTD}, referred to as the dome copy, was docked above the other four NUP358^{NTD} copies and the distal NUP93^{SOL}, with its N and C termini oriented toward the C termini of the outer distal and inner proximal copies of NUP358^{NTD}, respectively (Fig. 4E). Though unexpected, the placement of the dome NUP358^{NTD} was the second most confident docking solution into both the current ~12-Å and previously reported ~23-Å cryo-ET maps of the intact human NPC (*44*) (fig. S57). The placement of 40 molecules of NUP358 per NPC is in agreement with previous experimental lower-bound stoichiometry estimates of 32 molecules of NUP358 (*86*). Finally, we successfully placed the composite structure of the entire cytoplasmic outer-ring protomer, including all five NUP358^{NTD} copies, into an anisotropic ~7-Å region of a composite single-particle cryo-EM map of the *X. laevis* NPC cytoplasmic outer-ring protomer (figs. S58 and S59) (*45*).

The arrangement of five NUP358^{NTD} copies in each spoke places their C termini in proximity of each other, projecting the remaining domains toward the cytoplasm. Consequently, the oligomerization domains of the five NUP358^{NTD} copies are constrained to form a homomeric assembly within the same spoke (Fig. 4, F and G). The oligomerization of NUP358 observed in the NUP358^{OE} crystal structure and SEC-MALS analysis would boost the avidity of NUP358 attachment to the cytoplasmic face of the NPC (Figs. 3, G and H, and 4G).

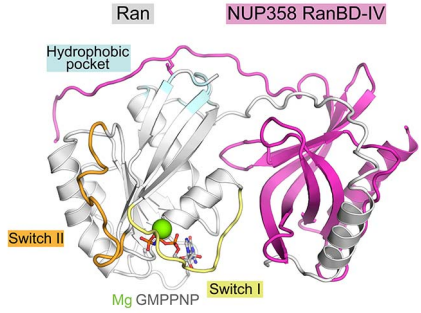
NUP358 is dispensable for NPC integrity during interphase

Our quantitative docking showed that NUP358^{NTD} is the primary attachment point for NUP358 at the cytoplasmic face of the NPC. To validate this result physiologically, we sought to determine the subcellular localization of structure-guided NUP358 fragments in intact cells. To prevent default localization of ectopically expressed fragments at the nuclear envelope because of homo-oligomerization with the NUP358^{OE} of endogenous proteins, we generated an inducible NUP358-knockout cell line, in which an N-terminal auxin-inducible degron (AID) tag was inserted into both genomic *NUP358* loci (*AID*::*NUP358* HCT116) (fig. S60). Addition of auxin resulted in the rapid, selective, and complete degradation of endogenous NUP358 within 3 hours, confirmed by the loss of immunofluorescent nuclear envelope rim staining and Western blot analysis of cellular NUP358 protein levels (Fig. 5A and figs. S60 and S61).

To identify the minimal NUP358 region necessary and sufficient for nuclear envelope targeting, we generated a systematic series of hemagglutinin (HA)–tagged N- and C-terminal fragments, splitting the protein into two pieces after the NTD, OE, RanBD-I, or ZFD, and determined their subcellular localization by immunofluorescence microscopy. NUP358 targeting to the nuclear envelope in the absence of auxin required both NTD and



Movie 3. Comparison of NUP358 and Nup153 ZnF Ran(GDP) complexes. Crystal structures of the six NUP358 ZnF•Ran(GDP) and four NUP153 ZnF•Ran(GDP) cocrystal structures are shown individually followed by a superposition. A 360° rotation of the NUP358 and NUP153 ZnF superposition and Ran(GDP) is provided, with a zoom-in view showing hydrophobic residues of the ZnF N-terminal extension buried in the Ran hydrophobic pocket. Colors are as in Fig. 3K.



Movie 4. Comparison of NUP358 and Nup50 RanBD Ran(GMPPNP) complexes. A 360° rotation of the of NUP358$^{RanBD\text{-}IV}$•Ran(GMPPNP) cocrystal structure, colored as in Fig. 3L. A zoom-in view is provided, transitioning between the four different Nup358 RanBD structures and the single NUP50^{RanBD} structure, showing hydrophobic residues of the NUP358 RanBD N-terminal extension buried in the Ran hydrophobic pocket. Finally, the interaction of individual RanBD basic patches with the Ran acidic tail is shown, colored as in fig. S55D.

OE regions, whereas all other domains were dispensable (Fig. 5B and fig. S62). When NUP358^{NTD} or NUP358^{OE} were tested in isolation, neither was found to be sufficient, with both domains exhibiting strong nuclear staining (Fig. 5B). Introduction of the NUP358 2R5K mutation, located on the NUP358^{NTD} concave surface in contact with the CNC, either abolished or severely reduced nuclear envelope rim staining when introduced into HA-NUP358$^{NTD\text{-}OE}$ or HA-NUP358^{FL}, respectively (Fig. 5B). Analogously, NUP358 oligomerization is required for localization to the nuclear envelope, with introduction of the oligomerization-deficient LIQIML mutation eliminating nuclear envelope rim staining of both HA-NUP358$^{NTD\text{-}OE}$ and HA-NUP358^{FL} (Fig. 5B). Notably, we repeated the fluorescence microscopy analysis after auxin-induced depletion of endogenous NUP358 and obtained identical results (fig. S63).

The previous ~38-Å cryo-ET map of the human NPC showed loss of the distal cytoplasmic CNC ring and potentially other CF nups in the absence of NUP358, leading to the conclusion that NUP358 is required for the integrity of the interphase NPC (*44*). To determine whether the architectural stability of the interphase NPC depends on NUP358, we sought to analyze the effect of auxin-induced NUP358 depletion on the subcellular localization of eight nups representative of all NPC subcomplexes by immunofluorescence microscopy (Fig. 5A). Because previous studies showed NUP358 depletion results in cell-cycle arrest at the G_2-to-M phase transition (*87*), we first monitored the levels of cell-cycle markers to determine a cell-cycle length of ~14 hours, consistent with previous reports for HCT cells (fig. S64) (*88*). We then induced NUP358-degradation in nocodazole-synchronized cells and imaged nups at various time points before cells entered mitosis and at 24 hours. Although NUP358 nuclear envelope rim staining was rapidly lost 2 hours after induction of degradation and remained absent throughout the remaining time points, all eight representative nups continued to display robust nuclear envelope rim staining (Fig. 5A). This suggests that NPC integrity is not dependent on NUP358 attachment to the NPC and also demonstrates the specificity of the auxin-induced NUP358 knockout. To reconcile the apparent conflict between our results and the aforementioned cryo-ET study, we investigated whether NUP358 depletion led to release of nups from the nuclear fraction during cellular fractionation. Indeed, we observed auxin-dependent leakage of NUP214, NUP88, and NUP160 from the nuclear to the cytoplasmic fraction (fig. S61C). Curiously, we also consistently observed a reduction of the nuclear basket nup ELYS in the nuclear fraction upon NUP358 depletion.

Together, these data confirm the quantitative docking of NUP358^{NTD}, validate the physiological relevance of NUP358^{OE}-mediated bundling, and establish that NUP358 is dispensable for the architectural integrity of the assembled interphase NPC, although its depletion made the structural integrity of the cytoplasmic face of the NPC susceptible to the biochemical stresses inherent to cell fractionation. Future studies are needed to establish the extent of NUP358's role in the formation of the double–CNC ring architecture during NPC assembly.

NUP358 plays a general role in translation of exported mRNAs

Export of mRNA from the nucleus to the cytoplasm is an essential step in the expression of eukaryotic proteins (Fig. 5C) (*43*). Our biochemical analysis revealed that NUP358 has multiple RNA-binding domains distributed throughout the protein, suggesting a potential role in RNA export and mRNP remodeling (Fig. 2G). The docking of five NUP358^{NTD} copies in the intact human NPC revealed occlusion of the RNA/NUP88^{NTD}-binding surfaces of the dome and inner distal NUP358 copies by the CNC stalk but exposure of the remaining copies' binding sites (fig. S65). Thus, some NUP358^{NTD} copies could potentially be simultaneously attached to the NPC and dynamically interact with RNA/NUP88^{NTD}.

Previous studies had found that efficient translation of secretory proteins requires NUP358 binding to the ~63-nucleotide GC-rich signal sequence coding region (SSCR) of mRNAs encoding secretory proteins (*89*). NUP358 knockdown by short hairpin RNAs was shown to prevent the translation of various secretory protein reporters but had no effect on the distribution of mRNA in the cell (*89*). These experiments involved extended

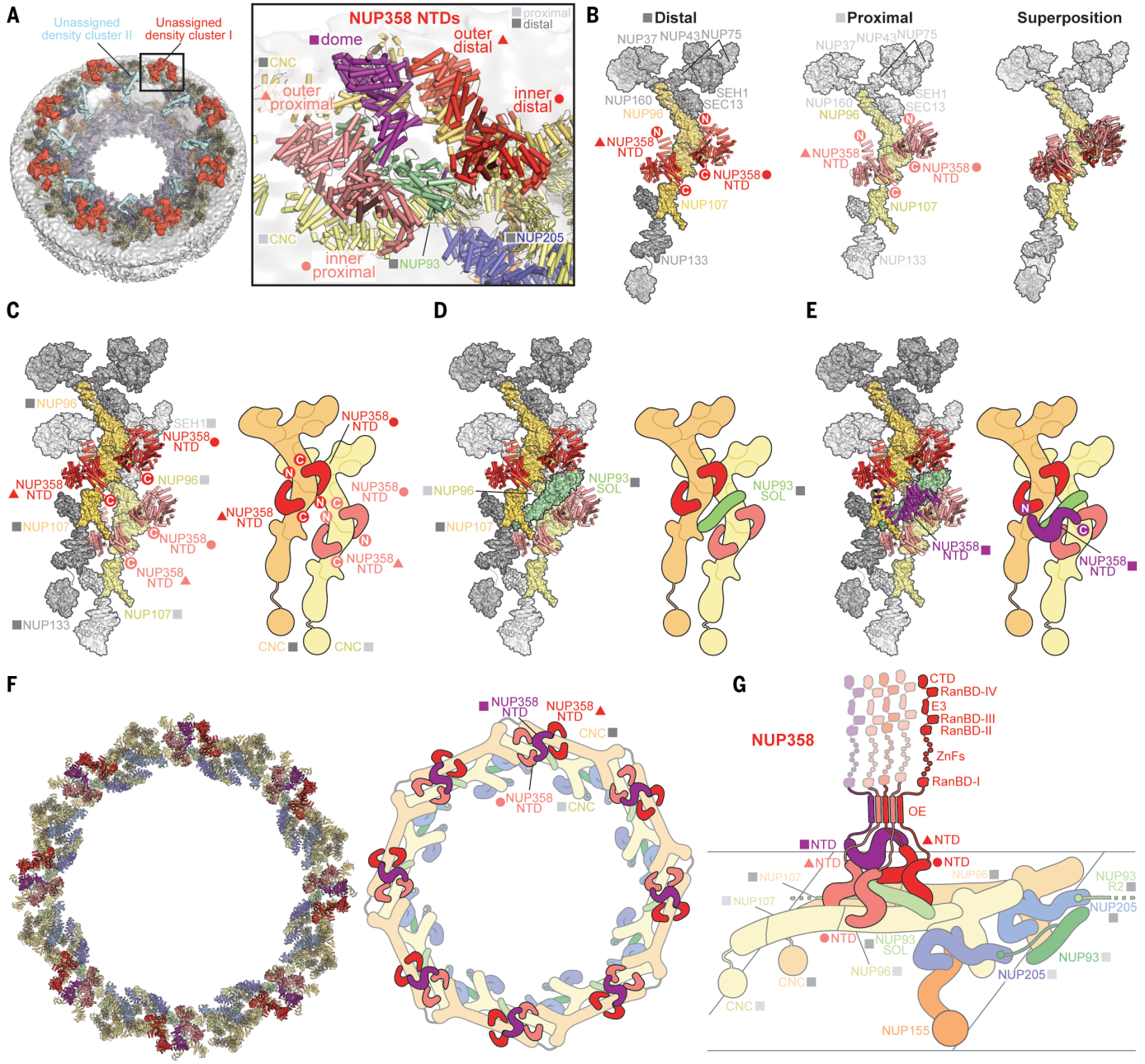


Fig. 4. Docking of NUP358^{NTD} on the cytoplasmic face of the NPC. (**A**) Overview of the NPC cytoplasmic face with isosurface rendering of unexplained density clusters I (red) and II (cyan) of the ~12-Å cryo-ET map of the intact human NPC. The inset indicates the location of the magnified view showing cartoon representations of five copies of NUP358^{NTD} docked in unassigned density cluster I. (**B**) Comparison of the binding of two NUP358^{NTD} copies (cartoon representation) to distal and proximal CNCs (surface representation). (**C** to **E**) Architecture of the pentameric NUP358^{NTD} bundle attachment site on a cytoplasmic outer-ring spoke with cartoon- and surface-represented structures (left) and schematics (right), sequentially illustrating the placement of (C) four copies of NUP358^{NTD} around NUP96 and NUP107 interfaces on the stalks of tandem-arranged Y-shaped CNCs; (D) a distal copy of NUP93^{SOL} collocated at the center of the NUP358^{NTD} bundle, interfacing with both proximal and distal CNC stalks; and (E) the NUP358^{NTD} dome copy interfacing with the stalk-attached NUP358^{NTD} quartet beneath. (**F**) Overview of the entire cytoplasmic face of the NPC in cartoon representation and as a schematic, illustrating the distribution of 40 NUP358^{NTD} copies anchored as pentameric bundles across the eight NPC spokes. (**G**) Schematic of NUP358 attachment to the cytoplasmic outer-ring spoke. The NUP358 pentameric bundle is linked together by interactions between OEs. Anchored by NUP358^{NTD}, the rest of the NUP358 domains are linked by unstructured linker sequences and are expected to freely project from the cytoplasmic face of the NPC. Distal and proximal positions are labeled according to the legend in (A).

incubation periods and achieved only a partial NUP358 knockdown, potentially allowing secondary phenotypes to emerge from prolonged NPC disruption, non–NPC-related NUP358 effects, or defective postmitotic NPC reassembly.

With the ability to rapidly deplete NUP358 in our AID cell line, we sought to examine whether NUP358 is directly involved in mRNA export or mRNP remodeling by monitoring the subcellular distribution of 5-ethynyl uridine (5-EU) pulse-labeled, newly synthesized RNA after NUP358 depletion. In situ–labeled RNA was visualized by fluorescence microscopy at hourly intervals during the chase for up to 6 hours (Fig. 5D). At early time points, 90

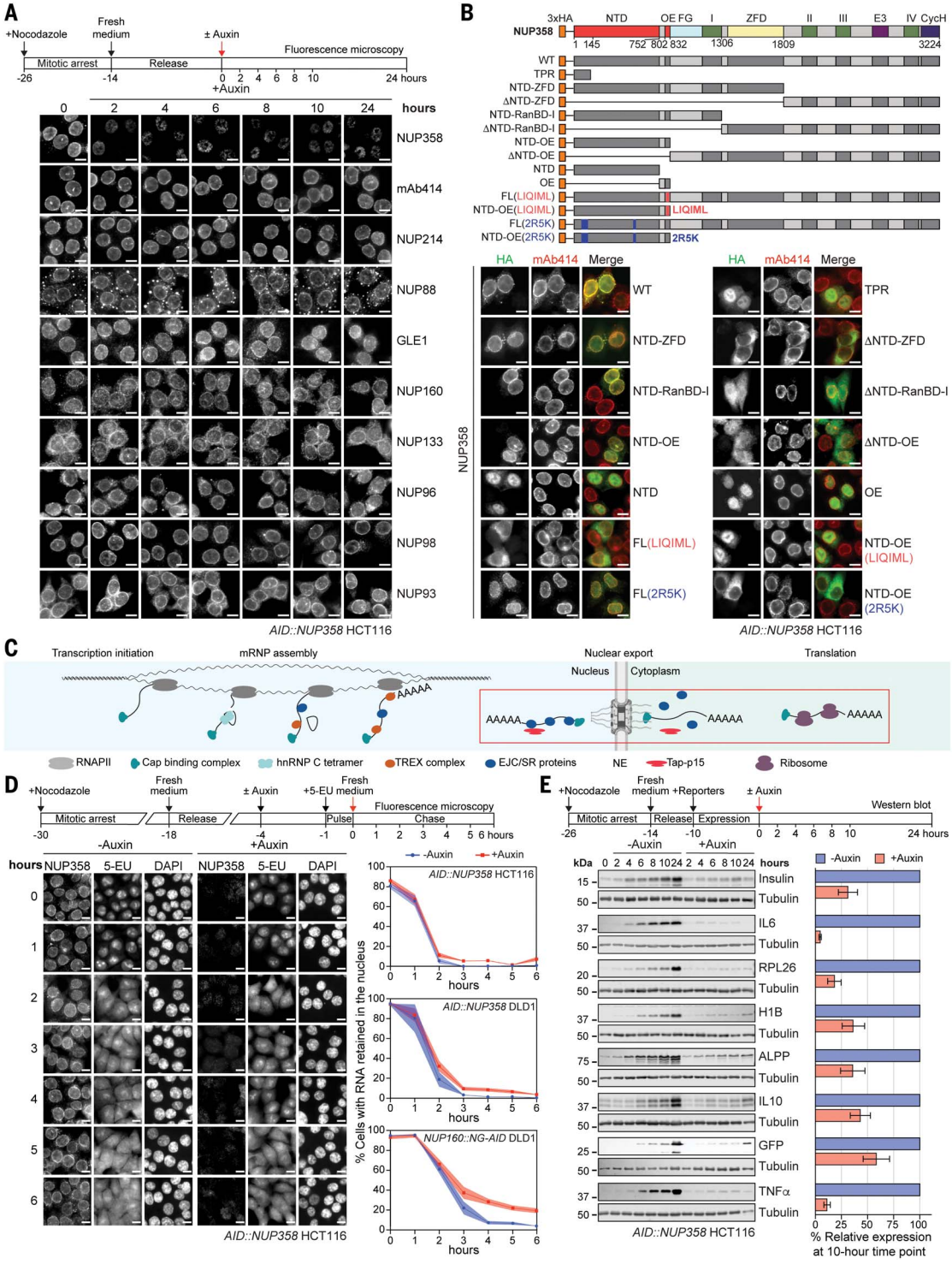


Fig. 5. NUP358 plays a general role in translation of exported mRNA. (**A**) Subcellular localization by immunofluorescent staining of endogenous nups in synchronized *AID*::*NUP358* HCT116 cells at the indicated time points after auxin-induced NUP358 depletion. Cytoplasmic puncta arise from NUP88 overexpression inherent to HCT116 cells. (**B**) Subcellular localization of N-terminally 3×HA-tagged NUP358 variants in *AID*::*NUP358* HCT116 cells by immunofluorescence microscopy. mAb414 staining marks the nuclear envelope rim location. The domain structure of the transfected NUP358 variants is shown (top). (**C**) Schematic illustrating the life cycle of mRNPs from transcription to maturation, export, remodeling at the cytoplasmic face of the NPC, and translation. Steps associated with the NPC are highlighted by the red box. (**D**) Time-resolved analysis of RNA nuclear retention in synchronized *AID*::*NUP358* HCT116 cells upon auxin-induced NUP358 depletion visualizing 5-EU metabolically pulse-labeled RNA. Representative images (left) and quantitation (right) of the proportion of cells (n > 200 per time point) with nuclear RNA retention are shown with mean values (squares) and the respective standard error (shaded area) of triplicate experiments. Quantitation from unsynchronized *AID*::*NUP358* DLD1 and *NUP160*::*NG-AID* DLD1 cells are also shown. (**E**) Time-resolved Western blot analysis of the expression of C-terminally 3×FLAG-tagged reporter proteins in synchronized *AID*::*NUP358* HCT116 cells upon auxin-induced NUP358 depletion. Quantitated reporter expression in auxin-treated cells was normalized to expression in control cells at the 10-hour time point. Experiments were performed in triplicate, with mean and associated standard error shown. Scale bars, 10 μm. Experimental timelines are shown above each experiment.

to 100% of cells displayed a strong nuclear 5-EU–labeled RNA signal that decreased over time with a concomitant increase in the cytoplasmic signal, indicative of RNA being exported. After 6 hours, only ~5% of NUP358-depleted cells exhibited nuclear retention of labeled RNA, compared with <1% of control cells (Fig. 5D). Because of this subtle effect, we sought to confirm the result in *AID*::*NUP358* DLD1 cells (Fig. 5D and figs. S66 and S67). Similar to the *AID*::*NUP358* HCT116 cell results, only ~3% of NUP358-depleted DLD1 cells exhibited nuclear RNA retention after a 6-hour chase (Fig. 5D and fig. S68). By contrast, nuclear RNA retention was present in ~20% of

NUP160::*NG-AID* DLD1 cells after depletion of NUP160, the knockout of which causes RNA retention in *S. cerevisiae*, demonstrating the principle suitability of our experimental approach (Fig. 5D and fig. S68) (*90*).

Next, we analyzed the fate of the genetic message downstream of mRNA export by examining the dependence of cellular protein expression on NUP358 in *AID*::*NUP358* HCT116 cells using eight different reporter constructs. Synchronized cells were transfected with C-terminally FLAG-tagged reporter constructs before NUP358 depletion, and the amount of reporter in whole-cell extracts was determined by Western blot analysis (Fig. 5E and fig. S69). First, we focused our analysis on representative secretory protein reporter constructs including insulin, interleukin-10 (IL-10), IL-6, tumor necrosis factor–α (TNFα), and membrane-bound placental alkaline phosphatase (ALPP). NUP358 depletion significantly reduced the cellular levels of all five reporters. We also tested nonsecretory protein reporters and, contrary to the previous observation of secretory protein-effect specificity, found that NUP358 depletion also significantly reduced the cellular levels of ribosomal protein L26 (RPL26), green fluorescence protein (GFP), and histone 1B (H1B) reporters (Fig. 5E and fig. S69).

In summary, our data confirm that NUP358 depletion does not result in marked nuclear RNA accumulation, but it nevertheless affects the efficient translation of secreted and membrane-bound proteins, as previously proposed (*89*). However, our findings also demonstrate that the observed translational defect is not restricted to secretory proteins, which suggests a more general role of NUP358 in mRNP-remodeling events that occur at the cytoplasmic face of the NPC after mRNA export.

Characterization of NUP358 harboring ANE1 mutations

Acute necrotizing encephalopathy (ANE) is an autoimmune disease in which previously healthy children experience a cytokine storm after common viral infections, resulting in brain inflammation and rapid deterioration from seizures to coma that can ultimately be fatal (*91*). ANE1, the familial and recurring form of ANE, has been associated with four distinct NUP358 mutations: T585M, T653I, I656V, and W681C (*91*, *92*). All four ANE1 mutations map to the C-terminal α-helical solenoid of $NUP358^{NTD}$ (fig. S70). Apart from T585, which is exposed on the surface, the ANE1 mutations locate in the closely packed hydrophobic core (fig. S70C). We determined cocrystal structures of $NUP358^{NTD}$ harboring the individual ANE1 mutations T585M, T653I, and I656V in complex with sAB-14, revealing no substantial structural changes, with root mean square deviation calculated over 746 C_α atoms of ~0.5 Å (fig. S70 and table S10). Moreover, we did not detect differences in nuclear envelope rim staining or binding to $NUP88^{NTD}$ or ALPP SSCR RNA (fig. S71). Notably, we found that in vitro thermosolubility of the W681C, T653I, and I656V mutants was reduced at temperatures below body temperature (fig. S72) but increased beyond wild-type levels by binding to sAB-14 in all three mutants (fig. S73).

Together, our results indicate that ANE1 mutations neither directly disturb the fold observed in the crystal structure nor affect the known cellular functions of NUP358. Our observation of a substantially reduced thermosolubility of NUP358 ANE1 mutants is notable, considering that the sudden onset of symptoms appears to require a fever-inducing trigger such as a viral infection (*91*). Future studies will be needed to systematically assess whether ANE1 mutations affect unknown cellular functions of NUP358.

Structural and biochemical analysis of $NUP88^{NTD}$•$NUP98^{APD}$•$NUP214^{TAIL}$

Besides NUP358, the $NUP88^{NTD}$•$NUP98^{APD}$•$NUP214^{TAIL}$ complex had up to now been another CF component for which atomic-level structural information had remained unavailable. Through extensive screening of crystallization fragments and conditions, we solved the structure of the heterodimeric $NUP88^{NTD}$•$NUP98^{APD}$ at 2.0-Å resolution (Fig. 6H and table S17). Despite low sequence homology, the overall architecture of the $NUP88^{NTD}$•$NUP98^{APD}$ complex is conserved from fungi to humans, although the orientation of $NUP98^{APD}$ relative to $NUP88^{NTD}$ varies between the cocrystal structures of human and fungal orthologs by as much as ~20° (figs. S74 to S77 and Movie 5) (*11*, *59*). For a detailed description of the structure, see the supplementary text (figs. S76 to S81).

Because the $NUP214^{TAIL}$-$NUP88^{NTD}$ interaction was crystallographically intractable, we mapped a minimal $NUP88^{NTD}$-binding region spanning NUP214 residues 938 to 955 by systematic truncation (figs. S78 and S79). $NUP214^{TAIL}$ forms a hydrophobic interaction with $NUP88^{NTD}$ at the 6CD insertion, which was abolished by a combined $NUP88^{NTD}$ LLL mutation, analogous to a mutation we had previously shown to abolish the interaction between the *S. cerevisiae* orthologs $Nup159^{TAIL}$ and $Nup82^{NTD}$ (fig. S79) (*59*). Notably, this $NUP88^{NTD}$ LLL mutation straddles a naturally occurring D434Y mutation in NUP88 that is linked to a fatal disorder called fetal akinesia deformation sequence, which is associated with congenital malformations and impaired fetal movement (fig. S80) (*93*). Given its location, the D434Y mutation is expected to interfere with the $NUP214^{TAIL}$ interaction.

Combined, our structural and biochemical analysis of NUP88, NUP214, NUP98, and their interactions shows that their shape, mode of interaction, and the overall architecture of their complexes are evolutionarily conserved from fungi to humans, despite primary sequence divergence.

Docking of the CFNC into unassigned cluster II

After the placement of five $NUP358^{NTD}$ copies into unassigned density cluster I, we reasoned that the remaining unassigned density cluster II would represent the CFNC. Unassigned density cluster II is composed of two near-perpendicular tube-like segments that bisect the NUP75 arms of the distal and proximal Y-shaped CNCs—a globular segment lodged between the base of the long tube-like segment and the proximal NUP75 arm, and a dumbbell-shaped globular density that projects toward the central transport channel (Fig. 6A). Owing to the small size and lack of distinctive shape features, the quantitative docking of $NUP88^{NTD}$•$NUP98^{APD}$, $NUP214^{NTD}$•DDX19, $GLE1^{CTD}$•$NUP42^{GBM}$, $GLE1^{CTD}$•$NUP42^{GBM}$•DDX19, RAE1•$NUP98^{GLEBS}$, $NUP358^{RanBD}$•Ran, $NUP358^{ZnF}$•Ran, and $NUP358^{CTD}$ into the ~12-Å cryo-ET map of the intact human NPC, from which all of the previously explained density had been subtracted, did not result in high-confidence solutions (fig. S82). We therefore took the less objective approach of manual placement based on shape complementarity and biochemical restraints, followed by local rigid-body refinement.

We used the *C. thermophilum* and *X. laevis* CNT crystal structures, the latter containing NUP62, as a template for the polyalanine model of the coiled-coil segments (CCSs) 1 and 2 of the CFNC hub (*10*, *11*). Notably, the CCS1 and CCS2 models based on CNT structures matched the shape and dimensions of two near-perpendicular segments of tube-like density, which suggests that the CFNC-hub coiled-coil architecture is similar to that of the CNT (Fig. 6, B and C, and fig. S83A). $NUP88^{NTD}$•$NUP98^{APD}$ fit best at the base of the long CCS1 segment, interfacing with the NUP75 arm of the proximal CNC. The tentative placement would be consistent with the biochemically mapped interaction between $NUP88^{NTD}$•$NUP98^{APD}$ and the $NUP214^{TAIL}$ segment expected to emanate from the C-terminal base of the CCS3 segment and thus restrain $NUP88^{NTD}$•$NUP98^{APD}$ near the CFNC-hub base (Fig. 6, B to E). A dumbbell-shaped density interfacing with $NUP88^{NTD}$•$NUP98^{APD}$ and extending toward the central transport channel is consistent with the shape and size of the $NUP214^{NTD}$•DDX19 crystal structure, although it could also be explained by other more transiently tethered components of the nucleocytoplasmic transport machinery or

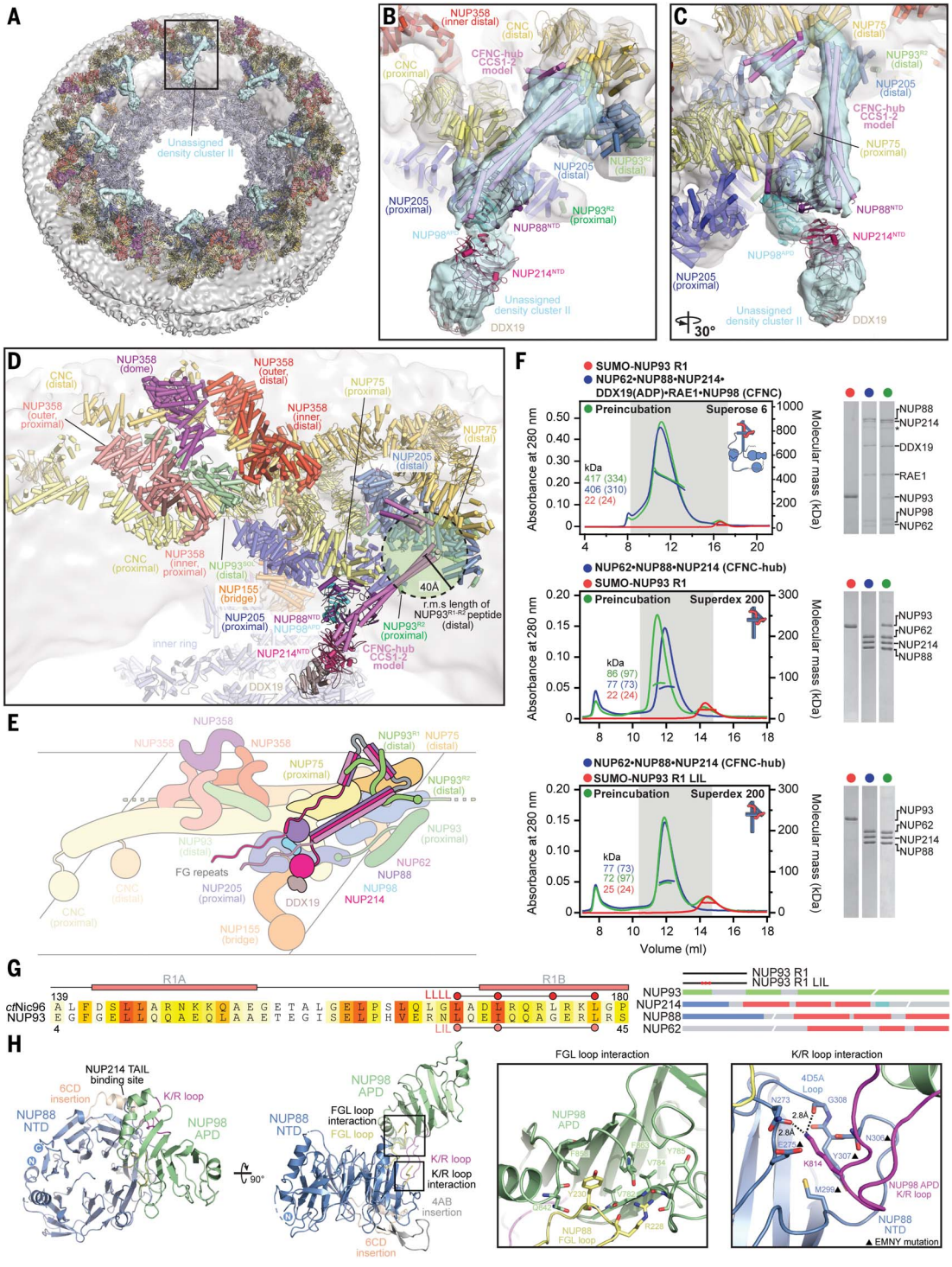
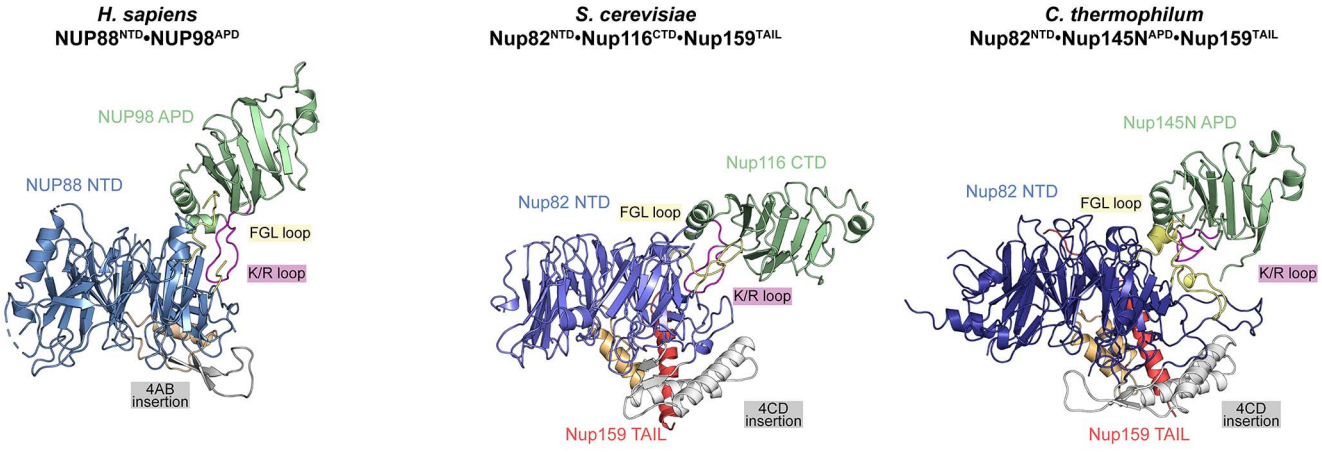


Fig. 6. Docking analysis reveals that NUP93 anchors the CFNC to the cytoplasmic outer ring. (**A**) Overview of the NPC cytoplasmic face with isosurface rendering of unexplained density cluster II (cyan) of the ~12-Å cryo-ET map of the intact human NPC. The inset indicates the location of the magnified view in (B). (**B** and **C**) Two views of manually placed poly-alanine models of CFNC-hub segments CCS1 and CCS2, as well as of the tentatively placed NUP88^{NTD}•NUP98^{APD} and NUP214^{NTD}•DDX19 cocrystal structures shown in cartoon representation. (**D**) Cartoon representation of a cytoplasmic face spoke. The root mean square (r.m.s.) end-to-end length estimate for the NUP93$^{R1\text{-}R2}$ linker defines a ~40-Å radius for the expected location of the distal NUP93^{R1} region. (**E**) Schematic of a cytoplasmic face spoke illustrating CFNC-hub anchoring by the distal NUP93^{R1} positioned by the distal NUP205-bound NUP93^{R2}. (**F**) SEC-MALS interaction analysis showing the binding of SUMO-NUP93^{R1} to the heterohexameric CFNC or CFNC hub and illustrating the abolished binding of the SUMO-NUP93^{R1} LIL mutant to the CFNC hub. Proteins were visualized by Coomassie brilliant blue staining. (**G**) (Left) An alignment of *C. thermophilum* Nic96 and *H. sapiens* NUP93 sequences shows the conservation of residues targeted by the LLLL and LIL mutations that abolish binding to the CNT. Residues are colored, according to a multispecies sequence alignment, from white (less than 55% similarity), to yellow (55% similarity), to red (100% identity), using the BLOSUM62 weighting algorithm. (Right) Domain architectures of NUP214, NUP88, NUP62, and NUP93. The location of the NUP93 LIL mutation is indicated (red dots). 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. (**H**) Two views of the NUP88^{NTD}•NUP98^{APD} cocrystal structure in cartoon representation. Inset boxes indicate regions of magnified view (right), of the NUP88^{NTD} FGL-loop interaction with NUP98^{APD} and of the NUP98^{APD} K/R-loop interaction with NUP88^{NTD}. Black triangles indicate alanine substitutions in the NUP88^{NTD} EMNY mutant.

cargo (Fig. 6, B and C, and fig. S83A). However, the NUP214^{NTD}•DDX19 complex forms tighter interactions when DDX19 is in its ADP-bound state and would therefore be expected to exist in the adenosine triphosphate (ATP)–depleted environment of the purified nuclear envelope (*50*, *63*). The placements of the CFNC-hub model and NUP88^{NTD}•NUP98^{APD} were further supported by manual docking into an anisotropic ~8-Å region of a composite single-particle cryo-EM map of the *X. laevis* cytoplasmic outer-ring protomer, although the map masking excluded the region that includes the dumbbell-shaped density (fig. S83B) (*45*).



Movie 5. Evolutionary conservation of the NUP88^{NTD}•NUP98^{APD} architecture. A 360° rotation of the NUP88^{NTD}•NUP98^{APD} cocrystal structure and the previously determined crystal structures of *S. cerevisiae* Nup82^{NTD}•Nup116^{CTD}•Nup159^{TAIL} (PDB ID 3PBP) (*59*) and *C. thermophilum* Nup82^{NTD}•Nup145N^{APD}•Nup159^{TAIL} (PDB ID 5CWW) (*11*), colored as in fig. S77.

Placement of the CFNC hub into the tube-like density puts CCS1, CCS2, and likely the unresolved CCS3 within reach of the ~40-Å root mean square length of the linker that tethers NUP93^{R1} to the NUP205-bound NUP93^{R2} (Fig. 6D and fig. S84). In the accompanying paper, we demonstrated that the NUP93^{R1} fragment (residues 2 to 93), like the orthologous *C. thermophilum* Nic96^{R1} assembly sensor, binds to the CNT complex of the inner ring (*42*). The proximity of the CFNC-hub coiled-coil segments to the expected NUP93^{R1} location suggested that NUP93^{R1} might act as assembly sensor for the CFNC hub as well. Indeed, the NUP93^{R1} fragment formed stable complexes with the intact CFNC and CFNC hub (Fig. 6F and fig. S85). Notably, the NUP93^{R1} LIL mutation that abolished CNT binding (*42*) also abolished the interaction with the CFNC hub (Fig. 6, F and G). To ensure that we did not miss an interaction of the *C. thermophilum* CFNC (*ct*CFNC), we evaluated whether Nic96^{R1} could bind the *ct*CFNC hub. In fact, Nic96^{R1} did not bind to the *ct*CFNC hub, consistent with our reconstitution results that identified two distinct assembly sensors for the *ct*CFNC in the CNC (fig. S86). These data indicate that the long-elusive assembly sensor anchoring point of the human CFNC is not provided by the Y-shaped CNC, but rather by the NUP205-positioned NUP93^{R1}, corroborated by the recent finding that NUP93 depletion displaces the CFNC nups NUP214, NUP88, and NUP62 from the nuclear envelope (*94*).

A second NUP93^{R1} assembly sensor emanating from the proximal NUP205-positioned NUP93^{R2} represents a potential anchoring site for a second, flexibly attached proximal CFNC (fig. S84). The placement of 16 copies of the CFNC on the cytoplasmic face of the NPC, half of which are unresolved in the ~12-Å cryo-ET map, is consistent with the previously established stoichiometry (*86*). Recent in situ ~37- and ~34-Å cryo-ET maps of the dilated human NPC (*95*, *96*) present unexplained elongated density near the expected location of the proximal NUP93^{R1} but could not be further interpreted at the current solutions (fig. S87).

Finally, we tentatively placed the human CF nup GLE1^{CTD}•NUP42^{GBM} crystal structure into a region of unexplained density in the ~12-Å cryo-ET map of the intact human NPC between the cytoplasmic bridge NUP155 and the cytoplasmic face of the nuclear envelope, consistent with our previous analysis (fig. S88) (*63*).

Steric occlusion is insufficient to explain asymmetric decoration of the NPC

Having assigned all cytoplasmic density of clusters I and II to NUP358 pentameric bundles and CFNCs, respectively, we next evaluated whether any structural features prevent NUP358 or CFNC mislocalization at the nuclear face of the NPC. We found that unexplained nuclear density adjacent to the NUP160 arms of the Y-shaped CNCs could be assigned to 16 copies of the structured N-terminal domains of the nuclear basket nup ELYS (fig. S89) (*25*). The ELYS domains did not overlap with nuclear regions equivalent to the sites occupied by NUP358 and CFNC on the cytoplasmic face, thereby excluding that steric competition with NUP358 or CFNC prevents ELYS mislocalization (Fig. 7, A and B). On the contrary, the ~12-Å cryo-ET map revealed rod-shaped unassigned densities atop the nuclear outer ring in regions equivalent to NUP358 sites on the cytoplasmic face (Fig. 7C). Analogously, we examined whether recruitment of the CFNC to the nuclear face was prevented by steric hindrance from a nuclear basket component. Although the NUP205-NUP93^{R2} attachment site from which NUP93^{R1} is flexibly projected remains unencumbered, an unassigned rod-shaped cryo-ET density present on the nuclear face overlaps with an area equivalent to the CFNC-hub docking sites on the cytoplasmic face (Fig. 7C). Together, these findings suggest that mechanisms other than steric competition alone, such as active nuclear transport of asymmetric nups, as previously indicated for NUP214 and NUP153 (*81*, *97*), are key determinants of the asymmetric localization of NUP358, CFNC, and ELYS.

Together, our data complete the near-atomic composite structure of the symmetric and cytoplasmic asymmetric portions of the human NPC (Fig. 8 and Movie 6).

Conclusions

Situated on the cytoplasmic face of the NPC, CF nups remodel mRNPs as they emerge from the central transport channel, ensuring directional transport of mRNA and preparing it for downstream translation. Given this essential cellular function, it is unsurprising that CF nups are a hotspot for mutations associated with currently incurable diseases, ranging from neurodegenerative and autoimmune disorders to aggressive cancers. Through a comprehensive analysis combining in vitro complex reconstitution, crystal structure determination, quantitative docking, and in vivo validation, we established a near-atomic composite structure of the cytoplasmic face of the human NPC.

Our biochemical reconstitution highlights the evolutionary conservation of the CFNC modular assembly, which consists of a central heterotrimeric coiled-coil hub that tethers two separate mRNP-remodeling complexes together. Despite the divergence in attachment mechanisms, the anchoring of two copies of the CFNC module to each of the eight NPC spokes appears to be an evolutionarily conserved

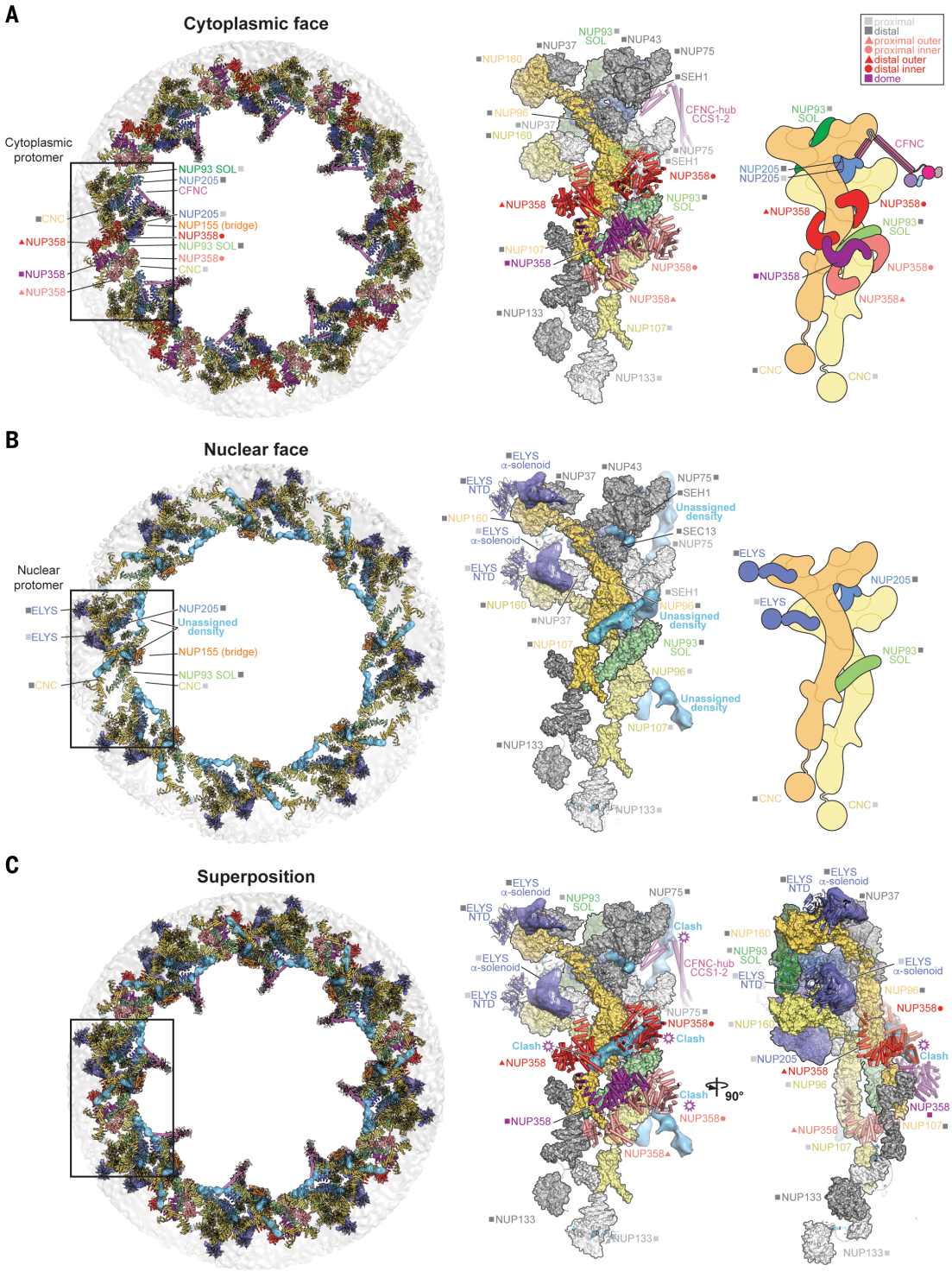


Fig. 7. Comparison of cytoplasmic and nuclear faces of the human NPC. (**A** and **B**) Overall top view (left); single-spoke protomer with symmetric core nups in surface, docked asymmetric nups in cartoon, and unexplained density of the ~12-Å cryo-ET map in isosurface representation (middle); and schematic (right) of the (A) cytoplasmic and (B) nuclear face of the intact human NPC. (**C**) Superpositions of the overall view (left) and two orthogonal views of single-spoke protomers (middle and right) of the nuclear and cytoplasmic faces, with hypothetical steric clashes between the CFs in cartoon representation, and the unassigned asymmetric nuclear density (cyan) indicated. Distal and proximal positions are labeled according to the legend. Inset boxes indicate regions of magnified protomer views (right).

architectural outcome: The *C. thermophilum* NPC presents two distinct assembly sensor motifs for the CFNC hub in the Nup37 and Nup145C subunits of each CNC. The human NPC reuses the NUP93 sensor for the assembly and anchoring of the CNT in the inner ring as an anchor for the CFNC in the cytoplasmic outer ring by intercalating two NUP205•NUP93 copies among the tandem-arranged CNCs of each spoke. Previous studies have also shown that the *S. cerevisiae* NPC incorporates a P-shaped CFNC dimer (*61*) to a single site within each of the eight outer-ring spokes (*40*, *41*).

In addition to the CFNC, the asymmetric CF nup decoration of the human NPC cytoplasmic face includes NUP358. Conjoined by an oligomerization element, pentameric bundles of the distinctively folded NUP358^{NTD} envelop the tandem-arrayed stalks of a CNC pair in each of the eight spokes. Each attached NUP358^{NTD} anchors an extensive ~2400-residue C-terminal region that harbors 14 different domains connected by unstructured linkers, thereby extending as much as ~60 nm from the outer ring (*98*). Our placement of the CNC and CF nups explains the entirety of the observed cytoplasmic

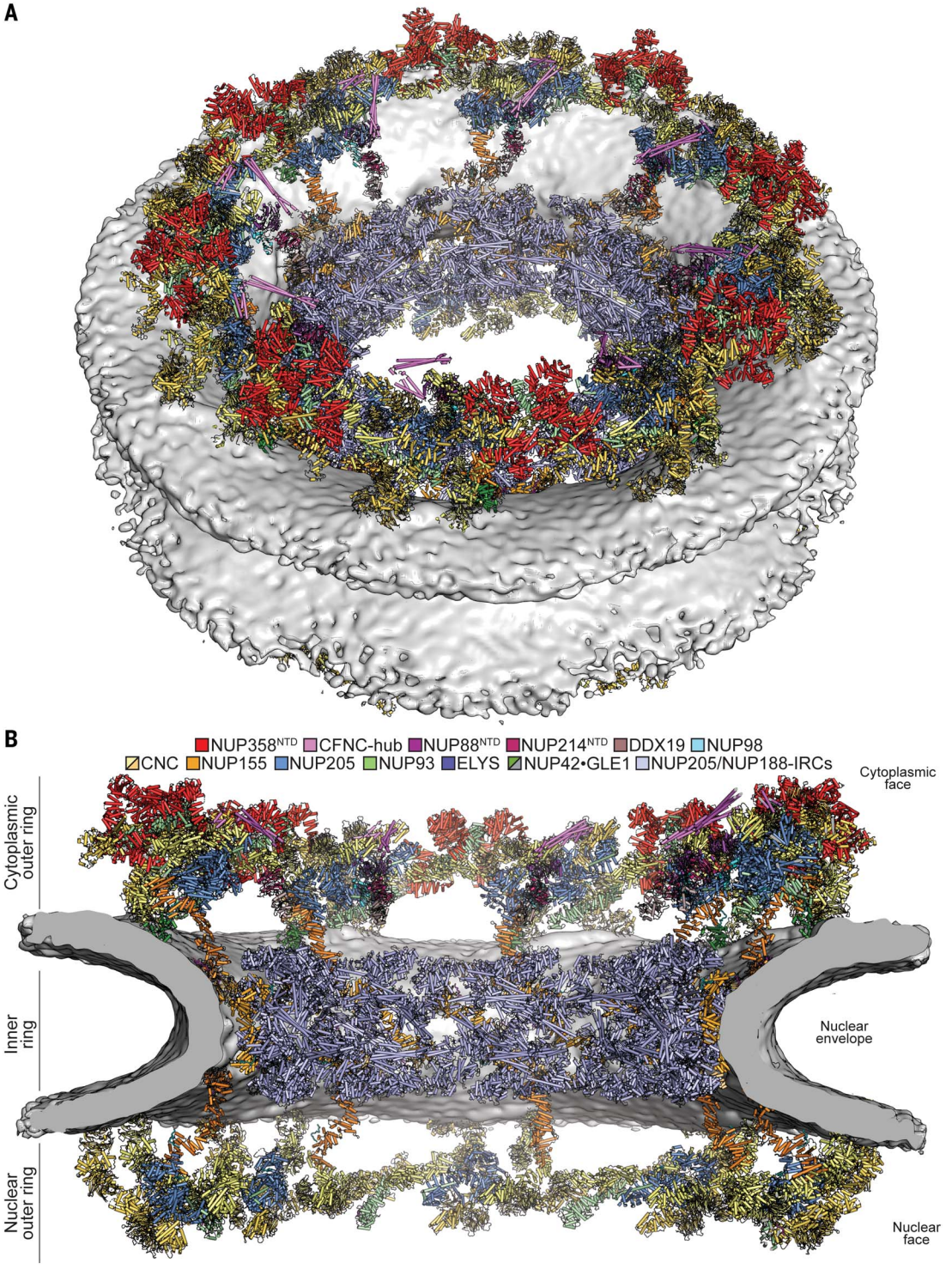


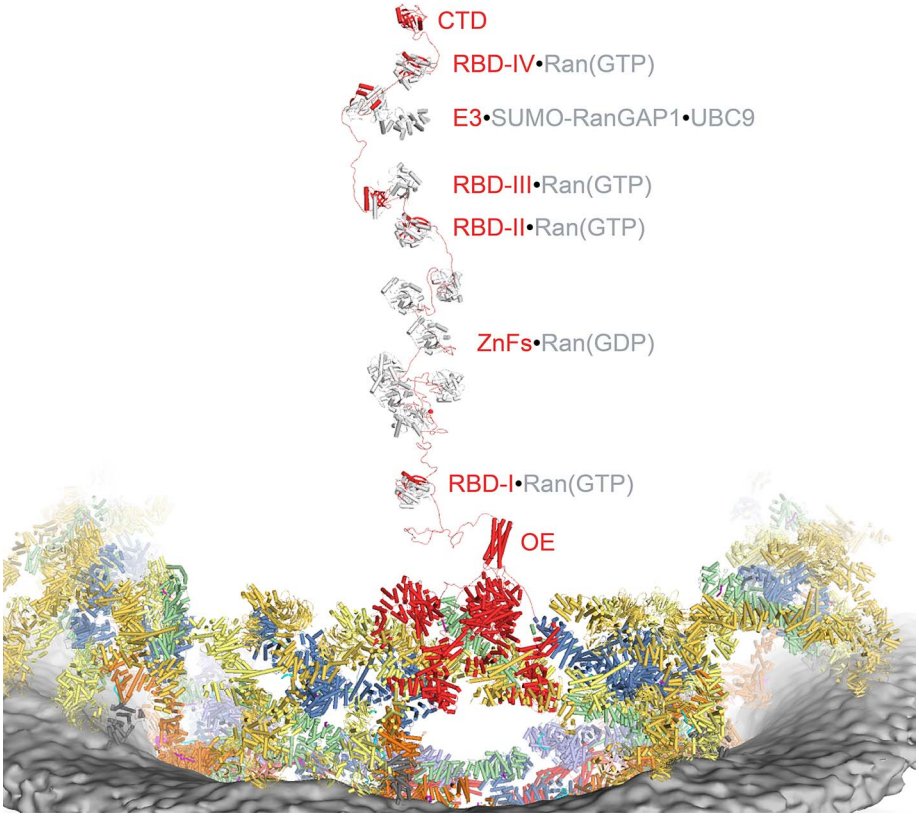
Fig. 8. Architecture of the human NPC cytoplasmic face. Near-atomic composite structure of the human NPC generated by docking individual nup and nup complex crystal and cryo-EM structures into a ~12-Å cryo-ET map of the intact human NPC, viewed from (**A**) the cytoplasmic face and (**B**) the central transport channel as a cross section. Newly placed structures include the quantitatively docked NUP358^{NTD} and the manually docked NUP88^{NTD}•NUP98^{APD}, NUP214^{NTD}•DDX19, GLE1^{CTD}•NUP42^{GBM}, ELYS^{NTD}, and a CFNC-hub model. The nuclear envelope is rendered as a gray isosurface. Nups are shown in cartoon representation and colored according to the legend.

face cryo-ET density, accounting for ~23 MDa of structured mass. The more flexibly attached regions of the CF nups that are not captured by the current subtomogram-averaged cryo-ET map account for an additional ~19 MDa of mass.

In addition to these flexibly attached structured domains, NUP358, NUP214, NUP98, and NUP42 contain extended FG-repeat regions emanating from various anchor points at the outer ring. The degree to which these regions contribute to the architectural integrity of the human NPC, as has been shown for the *S. cerevisiae* NPC (*99*), and the NPC's diffusion barrier remain important questions for future research.

We found that the interactions between the *C. thermophilum* CF nups and the CNC are modulated by the small molecule IP_6, the presence of which is required for mRNA export. Future studies are needed to address the concerted role of posttranslational modifications, second messengers and other small molecules, and macromolecular factors in regulating the assembly and functions of the NPC cytoplasmic face in mRNA export.

The integral membrane proteins and nuclear basket portions of the NPC represent an



Movie 6. Overview of the composite structure of the human NPC cytoplasmic face. The structures are shown docked into the cryo-ET reconstruction of the intact human NPC, colored according to Fig. 8. A pentameric bundle of NUP358^{NTD} is docked followed by relative placement of NUP358^{OE} and additional NUP358 domains, followed by the CFNC components. An overview of a single-spoke cytoplasmic face protomer is shown followed by a comparison cytoplasmic and nuclear face protomers.

outstanding challenge for structural determination. Nevertheless, our analysis has already identified that competition for binding sites could play a role in the segregation of CF and nuclear basket nups to opposite faces of the NPC. However, steric occlusion alone is insufficient to deterministically establish NPC polarity, whereby the correct asymmetric nups are segregated to either the cytoplasmic or the nuclear face, or the proximal NUP93 and NUP205 copies are excluded from the nuclear outer ring. Nuclear and cytoplasmic eviction mediated by the nucleocytoplasmic transport machinery is perhaps the most obvious candidate for a mechanism that maintains the polar subcellular segregation of asymmetric nups.

The data presented here provide a comprehensive biochemical foundation and a structural framework for the design of future experiments aimed at elucidating the multiple mechanistic steps involved in mRNP export and remodeling. This mechanistic insight will be vital for illuminating disease mechanisms associated with CF nup genetic variants and mechanisms by which viral virulence factors—e.g., SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) ORF6—hijack the functions of the NPC (*100*).

Our results represent a substantial step toward complete in vitro reconstitution of the NPC and establish a near-atomic composite structure of the entire cytoplasmic face of the human NPC. More broadly, they illustrate the effectiveness of our divide-and-conquer approach in successfully elucidating the near-atomic architecture of an assembly as large and complex as the NPC, serving as a paradigm for studying similar macromolecular machines, which remains a major frontier in structural cell biology.

Materials and methods summary

Full details of the materials and methods are presented in the supplementary materials. Briefly, the sources of materials and reagents are summarized in table S1. Bacterial, insect, and mammalian cell expression constructs and conditions are described in tables S2 to S4. Proteins were purified using standard chromatography techniques, and purification procedures are summarized in table S5. Purified proteins and complex formation were characterized by analytical SEC-MALS, summarized in table S6. LLPS of purified protein mixtures was analyzed by centrifugal separation followed by SDS–polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie staining, and by fluorescence microscopy after N-terminal amino labeling with fluorescent dyes. Nup-RNA binding interactions were assayed by EMSAs employing either ^{32}P-labeled or unlabeled nucleic acid probes, visualized by autoradiography or SybrGold-staining, respectively. Structures were determined by x-ray crystallography, with crystallization conditions and x-ray diffraction data collection, processing, and refinement statistics summarized in tables S7 to S17. Quantitative docking was performed by randomly placing and scoring densities simulated from crystal structures into ~12- and ~23-Å cryo-ET maps of the intact human NPC (*44*, *46*). Experimental structures used to generate the near-atomic composite structure of the intact human NPC are inventoried in table S18. NUP358 localization, NPC integrity, RNA export, and reporter expression levels were assessed in auxin-inducible degron cell lines.

REFERENCES AND NOTES

1. A. Hoelz, E. W. Debler, G. Blobel, The structure of the nuclear pore complex. *Annu. Rev. Biochem.* **80**, 613–643 (2011). doi: 10.1146/annurev-biochem-060109-151030; pmid: 21495847
2. D. H. Lin, A. Hoelz, The Structure of the Nuclear Pore Complex (An Update). *Annu. Rev. Biochem.* **88**, 725–783 (2019). doi: 10.1146/annurev-biochem-062917-011901; pmid: 30883195
3. B. Hampoelz, A. Andres-Pons, P. Kastritis, M. Beck, Structure and Assembly of the Nuclear Pore Complex. *Annu. Rev. Biophys.* **48**, 515–536 (2019). doi: 10.1146/annurev-biophys-052118-115308; pmid: 30943044
4. K. E. Knockenhauer, T. U. Schwartz, The Nuclear Pore Complex as a Flexible and Dynamic Gate. *Cell* **164**, 1162–1171 (2016). doi: 10.1016/j.cell.2016.01.034; pmid: 26967283
5. M. L. Yarbrough, M. A. Mata, R. Sakthivel, B. M. Fontoura, Viral subversion of nucleocytoplasmic trafficking. *Traffic* **15**, 127–140 (2014). doi: 10.1111/tra.12137; pmid: 24289861
6. A. Köhler, E. Hurt, Gene regulation by nucleoporins and links to cancer. *Mol. Cell* **38**, 6–15 (2010). doi: 10.1016/j.molcel.2010.01.040; pmid: 20385085
7. A. Cook, F. Bono, M. Jinek, E. Conti, Structural biology of nucleocytoplasmic transport. *Annu. Rev. Biochem.* **76**, 647–671 (2007). doi: 10.1146/annurev.biochem.76.052705.161529; pmid: 17506639
8. R. A. Pumroy, G. Cingolani, Diversification of importin-α isoforms in cellular trafficking and disease states. *Biochem. J.* **466**, 13–28 (2015). doi: 10.1042/BJ20141186; pmid: 25656054
9. Y. M. Chook, K. E. Süel, Nuclear import by karyopherin-βs: Recognition and inhibition. *Biochim. Biophys. Acta* **1813**, 1593–1606 (2011). doi: 10.1016/j.bbamcr.2010.10.014; pmid: 21029754
10. H. Chug, S. Trakhanov, B. B. Hülsmann, T. Pleiner, D. Görlich, Crystal structure of the metazoan Nup62•Nup58•Nup54 nucleoporin complex. *Science* **350**, 106–110 (2015). doi: 10.1126/science.aac7420; pmid: 26292704
11. T. Stuwe *et al.*, Architecture of the fungal nuclear pore inner ring complex. *Science* **350**, 56–64 (2015). doi: 10.1126/science.aac9176; pmid: 26316600
12. J. Fischer, R. Teimer, S. Amlacher, R. Kunze, E. Hurt, Linker Nups connect the nuclear pore complex inner ring with the outer ring and transport channel. *Nat. Struct. Mol. Biol.* **22**, 774–781 (2015). doi: 10.1038/nsmb.3084; pmid: 26344569

13. A. E. Hodel *et al*., The three-dimensional structure of the autoproteolytic, nuclear pore-targeting domain of the human nucleoporin Nup98. *Mol. Cell* **10**, 347–358 (2002). doi: 10.1016/S1097-2765(02)00589-0; pmid: 12191480
14. I. C. Berke, T. Boehmer, G. Blobel, T. U. Schwartz, Structural and functional analysis of Nup133 domains reveals modular building blocks of the nuclear pore complex. *J. Cell Biol.* **167**, 591–597 (2004). doi: 10.1083/jcb.200408109; pmid: 15557116
15. K. C. Hsia, P. Stavropoulos, G. Blobel, A. Hoelz, Architecture of a coat for the nuclear pore membrane. *Cell* **131**, 1313–1326 (2007). doi: 10.1016/j.cell.2007.11.038; pmid: 18160040
16. S. G. Brohawn, N. C. Leksa, E. D. Spear, K. R. Rajashankar, T. U. Schwartz, Structural evidence for common ancestry of the nuclear pore complex and vesicle coats. *Science* **322**, 1369–1373 (2008). doi: 10.1126/science.1165886; pmid: 18974315
17. T. Boehmer, S. Jeudy, I. C. Berke, T. U. Schwartz, Structural and functional studies of Nup107/Nup133 interaction and its implications for the architecture of the nuclear pore complex. *Mol. Cell* **30**, 721–731 (2008). doi: 10.1016/j.molcel.2008.04.022; pmid: 18570875
18. E. W. Debler *et al*., A fence-like coat for the nuclear pore membrane. *Mol. Cell* **32**, 815–826 (2008). doi: 10.1016/j.molcel.2008.12.001; pmid: 19111661
19. N. C. Leksa, S. G. Brohawn, T. U. Schwartz, The structure of the scaffold nucleoporin Nup120 reveals a new and unexpected domain architecture. *Structure* **17**, 1082–1091 (2009). doi: 10.1016/j.str.2009.06.003; pmid: 19576787
20. J. R. R. Whittle, T. U. Schwartz, Architectural nucleoporins Nup157/170 and Nup133 are structurally related and descend from a second ancestral element. *J. Biol. Chem.* **284**, 28442–28452 (2009). doi: 10.1074/jbc.M109.023580; pmid: 19674973
21. H. S. Seo *et al*., Structural and functional analysis of Nup120 suggests ring formation of the Nup84 complex. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14281–14286 (2009). doi: 10.1073/pnas.0907453106; pmid: 19706512
22. S. G. Brohawn, T. U. Schwartz, Molecular architecture of the Nup84-Nup145C-Sec13 edge element in the nuclear pore complex lattice. *Nat. Struct. Mol. Biol.* **16**, 1173–1177 (2009). doi: 10.1038/nsmb.1713; pmid: 19855394
23. V. Nagy *et al*., Structure of a trimeric nucleoporin complex reveals alternate oligomerization states. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 17693–17698 (2009). doi: 10.1073/pnas.0909373106; pmid: 19805193
24. S. Bilokapic, T. U. Schwartz, Molecular basis for Nup37 and ELY5/ELYS recruitment to the nuclear pore complex. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15241–15246 (2012). doi: 10.1073/pnas.1205151109; pmid: 22955883
25. S. Bilokapic, T. U. Schwartz, Structural and functional studies of the 252 kDa nucleoporin ELYS reveal distinct roles for its three tethered domains. *Structure* **21**, 572–580 (2013). doi: 10.1016/j.str.2013.02.006; pmid: 23499022
26. C. Xu *et al*., Crystal structure of human nuclear pore complex component NUP43. *FEBS Lett.* **589**, 3247–3253 (2015). doi: 10.1016/j.febslet.2015.09.008; pmid: 26391640
27. T. Stuwe *et al*., Architecture of the nuclear pore complex coat. *Science* **347**, 1148–1152 (2015). doi: 10.1126/science.aaa4136; pmid: 25745173
28. K. Kelley, K. E. Knockenhauer, G. Kabachinski, T. U. Schwartz, Atomic structure of the Y complex of the nuclear pore. *Nat. Struct. Mol. Biol.* **22**, 425–431 (2015). doi: 10.1038/nsmb.2998; pmid: 25822992
29. A. Hoelz, J. S. Glavy, M. Beck, Toward the atomic structure of the nuclear pore complex: When top down meets bottom up. *Nat. Struct. Mol. Biol.* **23**, 624–630 (2016). doi: 10.1038/nsmb.3244; pmid: 27273515
30. K. H. Bui *et al*., Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* **155**, 1233–1243 (2013). doi: 10.1016/j.cell.2013.10.055; pmid: 24315095
31. N. Handa *et al*., The crystal structure of mouse Nup35 reveals atypical RNP motifs and novel homodimerization of the RRM domain. *J. Mol. Biol.* **363**, 114–124 (2006). doi: 10.1016/j.jmb.2006.07.089; pmid: 16962612
32. S. Jeudy, T. U. Schwartz, Crystal structure of nucleoporin Nic96 reveals a novel, intricate helical domain architecture. *J. Biol. Chem.* **282**, 34904–34912 (2007). doi: 10.1074/jbc.M705479200; pmid: 17897938
33. N. Schrader *et al*., Structural basis of the nic96 subcomplex organization in the nuclear pore channel. *Mol. Cell* **29**, 46–55 (2008). doi: 10.1016/j.molcel.2007.10.022; pmid: 18206968
34. H. S. Seo, B. J. Blus, N. Z. Jankovic, G. Blobel, Structure and nucleic acid binding activity of the nucleoporin Nup157. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16450–16455 (2013). doi: 10.1073/pnas.1316607110; pmid: 24062435
35. P. Sampathkumar *et al*., Structure, dynamics, evolution, and function of a major scaffold component in the nuclear pore complex. *Structure* **21**, 560–571 (2013). doi: 10.1016/j.str.2013.02.005; pmid: 23499021
36. K. R. Andersen *et al*., Scaffold nucleoporins Nup188 and Nup192 share structural and functional properties with nuclear transport receptors. *eLife* **2**, e00745 (2013). doi: 10.7554/eLife.00745; pmid: 23795296
37. T. Stuwe, D. H. Lin, L. N. Collins, E. Hurt, A. Hoelz, Evidence for an evolutionary relationship between the large adaptor nucleoporin Nup192 and karyopherins. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2530–2535 (2014). doi: 10.1073/pnas.1311081111; pmid: 24505056
38. D. H. Lin *et al*., Architecture of the symmetric core of the nuclear pore. *Science* **352**, aaf1015 (2016). doi: 10.1126/science.aaf1015; pmid: 27081075
39. J. Kosinski *et al*., Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352**, 363–365 (2016). doi: 10.1126/science.aaf0643; pmid: 27081072
40. S. J. Kim *et al*., Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475–482 (2018). doi: 10.1038/nature26003; pmid: 29539637
41. M. Allegretti *et al*., In-cell architecture of the nuclear pore and snapshots of its turnover. *Nature* **586**, 796–800 (2020). doi: 10.1038/s41586-020-2670-5; pmid: 32879490
42. S. Petrovic *et al*., Architecture of the linker-scaffold in the nuclear pore. *Science* **376**, eabm9798 (2022). doi: 10.1126/science.abm9798
43. M. Stewart, Polyadenylation and nuclear export of mRNAs. *J. Biol. Chem.* **294**, 2977–2987 (2019). doi: 10.1074/jbc.REV118.005594; pmid: 30683695
44. A. von Appen *et al*., In situ structural analysis of the human nuclear pore complex. *Nature* **526**, 140–143 (2015). doi: 10.1038/nature15381; pmid: 26416747
45. G. Huang *et al*., Structure of the cytoplasmic ring of the Xenopus laevis nuclear pore complex by cryo-electron microscopy single particle analysis. *Cell Res.* **30**, 520–531 (2020). doi: 10.1038/s41422-020-0319-4; pmid: 32376910
46. S. Mosalaganti *et al*., AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science* **376**, eabm9506 (2022). doi: 10.1126/science.abm9506
47. P. Grandi *et al*., A novel nuclear pore protein Nup82p which specifically binds to a fraction of Nsp1p. *J. Cell Biol.* **130**, 1263–1273 (1995). doi: 10.1083/jcb.130.6.1263; pmid: 7559750
48. S. M. Bailer *et al*., Nup116p and nup100p are interchangeable through a conserved motif which constitutes a docking site for the mRNA transport factor gle2p. *EMBO J.* **17**, 1107–1119 (1998). doi: 10.1093/emboj/17.4.1107; pmid: 9463388
49. N. Belgareh *et al*., Functional characterization of a Nup159p-containing nuclear pore subcomplex. *Mol. Biol. Cell* **9**, 3475–3492 (1998). doi: 10.1091/mbc.9.12.3475; pmid: 9843582
50. C. Schmitt *et al*., Dbp5, a DEAD-box protein required for mRNA export, is recruited to the cytoplasmic fibrils of nuclear pore complex via a conserved interaction with CAN/Nup159p. *EMBO J.* **18**, 4332–4347 (1999). doi: 10.1093/emboj/18.15.4332; pmid: 10428971
51. A. K. Ho *et al*., Assembly and preferential localization of Nup116p on the cytoplasmic face of the nuclear pore complex by interaction with Nup82p. *Mol. Cell. Biol.* **20**, 5736–5748 (2000). doi: 10.1128/MCB.20.15.5736-5748.2000; pmid: 10891509
52. C. S. Weirich, J. P. Erzberger, J. M. Berger, K. Weis, The N-terminal domain of Nup159 forms a beta-propeller that functions in mRNA export by tethering the helicase Dbp5 to the nuclear pore. *Mol. Cell* **16**, 749–760 (2004). doi: 10.1016/j.molcel.2004.10.032; pmid: 15574330
53. A. R. Alcázar-Román, E. J. Tran, S. Guo, S. R. Wente, Inositol hexakisphosphate and Gle1 activate the DEAD-box protein Dbp5 for nuclear mRNA export. *Nat. Cell Biol.* **8**, 711–716 (2006). doi: 10.1038/ncb1427; pmid: 16783363
54. J. Napetschnig, G. Blobel, A. Hoelz, Crystal structure of the N-terminal domain of the human protooncogene Nup214/CAN. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1783–1788 (2007). doi: 10.1073/pnas.0610828104; pmid: 17264208
55. J. Napetschnig *et al*., Structural and functional analysis of the interaction between the nucleoporin Nup214 and the DEAD-box helicase Ddx19. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3089–3094 (2009). doi: 10.1073/pnas.0813267106; pmid: 19208808
56. H. von Moeller, C. Basquin, E. Conti, The mRNA export protein DBP5 binds RNA and the cytoplasmic nucleoporin NUP214 in a mutually exclusive manner. *Nat. Struct. Mol. Biol.* **16**, 247–254 (2009). doi: 10.1038/nsmb.1561; pmid: 19219046
57. Y. Ren, H. S. Seo, G. Blobel, A. Hoelz, Structural and functional analysis of the interaction between the nucleoporin Nup98 and the mRNA export factor Rae1. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10406–10411 (2010). doi: 10.1073/pnas.1005389107; pmid: 20498086
58. B. Montpetit *et al*., A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export. *Nature* **472**, 238–242 (2011). doi: 10.1038/nature09862; pmid: 21441902
59. K. Yoshida, H. S. Seo, E. W. Debler, G. Blobel, A. Hoelz, Structural and functional analysis of an essential nucleoporin heterotrimer on the cytoplasmic face of the nuclear pore complex. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16571–16576 (2011). doi: 10.1073/pnas.1112846108; pmid: 21930948
60. T. Stuwe, L. Schada von Borzyskowski, A. M. Davenport, A. Hoelz, Molecular basis for the anchoring of proto-oncoprotein Nup98 to the cytoplasmic face of the nuclear pore complex. *J. Mol. Biol.* **419**, 330–346 (2012). doi: 10.1016/j.jmb.2012.03.024; pmid: 22480613
61. M. Gaik *et al*., Structural basis for assembly and function of the Nup82 complex in the nuclear pore scaffold. *J. Cell Biol.* **208**, 283–297 (2015). doi: 10.1083/jcb.201411003; pmid: 25646085
62. R. Teimer, J. Kosinski, A. von Appen, M. Beck, E. Hurt, A short linear motif in scaffold Nup145C connects Y-complex with pre-assembled outer ring Nup82 complex. *Nat. Commun.* **8**, 1107 (2017). doi: 10.1038/s41467-017-01160-9; pmid: 29062044
63. D. H. Lin *et al*., Structural and functional analysis of mRNA export regulation by the nuclear pore complex. *Nat. Commun.* **9**, 2319 (2018). doi: 10.1038/s41467-018-04459-3; pmid: 29899397
64. J. Fernandez-Martinez *et al*., Structure and Function of the Nuclear Pore Complex Cytoplasmic mRNA Export Platform. *Cell* **167**, 1215–1228.e25 (2016). doi: 10.1016/j.cell.2016.10.028; pmid: 27839866
65. C. S. Weirich *et al*., Activation of the DExD/H-box protein Dbp5 by the nuclear-pore protein Gle1 and its coactivator InsP6 is required for mRNA export. *Nat. Cell Biol.* **8**, 668–676 (2006). doi: 10.1038/ncb1424; pmid: 16783364
66. J. Enninga, D. E. Levy, G. Blobel, B. M. Fontoura, Role of nucleoporin induction in releasing an mRNA nuclear export block. *Science* **295**, 1523–1525 (2002). doi: 10.1126/science.1067861; pmid: 11809937
67. D. Forler *et al*., RanBP2/Nup358 provides a major binding site for NXF1-p15 dimers at the nuclear pore complex and functions in nuclear mRNA export. *Mol. Cell. Biol.* **24**, 1155–1167 (2004). doi: 10.1128/MCB.24.3.1155-1167.2004; pmid: 14729961
68. S. Hutten, R. H. Kehlenbach, Nup214 is required for CRM1-dependent nuclear protein export in vivo. *Mol. Cell. Biol.* **26**, 6772–6785 (2006). doi: 10.1128/MCB.00342-06; pmid: 16943420
69. M. Hamada *et al*., Ran-dependent docking of importin-beta to RanBP2/Nup358 filaments is essential for protein import and cell viability. *J. Cell Biol.* **194**, 597–612 (2011). doi: 10.1083/jcb.201102018; pmid: 21859863
70. B. Quan, H. S. Seo, G. Blobel, Y. Ren, Vesiculoviral matrix (M) protein occupies nucleic acid binding site at nucleoporin pair (Rae1 • Nup98). *Proc. Natl. Acad. Sci. U.S.A.* **111**, 9127–9132 (2014). doi: 10.1073/pnas.1409076111; pmid: 24927547
71. J. Wu, M. J. Matunis, D. Kraemer, G. Blobel, E. Coutavas, Nup358, a cytoplasmically exposed nucleoporin with peptide repeats, Ran-GTP binding sites, zinc fingers, a cyclophilin A homologous domain, and a leucine-rich region. *J. Biol. Chem.* **270**, 14209–14213 (1995). doi: 10.1074/jbc.270.23.14209; pmid: 7775481
72. N. Yokoyama *et al*., A giant nucleopore protein that binds Ran/TC4. *Nature* **376**, 184–188 (1995). doi: 10.1038/376184a0; pmid: 7603572
73. M. P. Rout *et al*., The yeast nuclear pore complex: Composition, architecture, and transport mechanism. *J. Cell Biol.* **148**,

635–652 (2000). doi: 10.1083/jcb.148.4.635; pmid: 10684247
74. J. M. Cronshaw, A. N. Krutchinsky, W. Zhang, B. T. Chait, M. J. Matunis, Proteomic analysis of the mammalian nuclear pore complex. *J. Cell Biol.* **158**, 915–927 (2002). doi: 10.1083/jcb.200206106; pmid: 12196509
75. J. Joseph, M. Dasso, The nucleoporin Nup358 associates with and regulates interphase microtubules. *FEBS Lett.* **582**, 190–196 (2008). doi: 10.1016/j.febslet.2007.11.087; pmid: 18070602
76. S. A. Kassube *et al*., Crystal structure of the N-terminal domain of Nup358/RanBP2. *J. Mol. Biol.* **423**, 752–765 (2012). doi: 10.1016/j.jmb.2012.08.026; pmid: 22959972
77. M. Paduch *et al*., Generating conformation-specific synthetic antibodies to trap proteins in selected functional states. *Methods* **60**, 3–14 (2013). doi: 10.1016/j.ymeth.2012.12.010; pmid: 23280336
78. L. Holm, Using Dali for Protein Structure Comparison. *Methods Mol. Biol.* **2112**, 29–42 (2020). doi: 10.1007/978-1-0716-0270-6_3; pmid: 32006276
79. B. Hampoelz *et al*., Nuclear Pores Assemble from Nucleoporin Condensates During Oogenesis. *Cell* **179**, 671–686.e17 (2019). doi: 10.1016/j.cell.2019.09.022; pmid: 31626769
80. N. R. Yaseen, G. Blobel, Two distinct classes of Ran-binding sites on the nucleoporin Nup-358. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5516–5521 (1999). doi: 10.1073/pnas.96.10.5516; pmid: 10318915
81. S. Nakielny, S. Shaikh, B. Burke, G. Dreyfuss, Nup153 is an M9-containing mobile nucleoporin with a novel Ran-binding domain. *EMBO J.* **18**, 1982–1995 (1999). doi: 10.1093/emboj/18.7.1982; pmid: 10202161
82. T. Guan *et al*., Nup50, a nucleoplasmically oriented nucleoporin with a role in nuclear protein export. *Mol. Cell. Biol.* **20**, 5619–5630 (2000). doi: 10.1128/MCB.20.15.5619-5630.2000; pmid: 10891499
83. M. M. Higa, S. L. Alam, W. I. Sundquist, K. S. Ullman, Molecular characterization of the Ran-binding zinc finger domain of Nup153. *J. Biol. Chem.* **282**, 17090–17100 (2007). doi: 10.1074/jbc.M702715200; pmid: 17426026
84. N. Schrader *et al*., The crystal structure of the Ran-Nup153ZnF2 complex: A general Ran docking site at the nuclear pore complex. *Structure* **16**, 1116–1125 (2008). doi: 10.1016/j.str.2008.03.014; pmid: 18611384
85. J. R. Partridge, T. U. Schwartz, Crystallographic and biochemical analysis of the Ran-binding zinc finger domain. *J. Mol. Biol.* **391**, 375–389 (2009). doi: 10.1016/j.jmb.2009.06.011; pmid: 19505478
86. A. Ori *et al*., Cell type-specific nuclear pores: A case in point for context-dependent stoichiometry of molecular machines. *Mol. Syst. Biol.* **9**, 648 (2013). doi: 10.1038/msb.2013.4; pmid: 23511206
87. D. Salina, P. Enarson, J. B. Rattner, B. Burke, Nup358 integrates nuclear envelope breakdown with kinetochore assembly. *J. Cell Biol.* **162**, 991–1001 (2003). doi: 10.1083/jcb.200304080; pmid: 12963708
88. P. D. Pereira *et al*., Quantification of cell cycle kinetics by EdU (5-ethynyl-2′-deoxyuridine)-coupled-fluorescence-intensity analysis. *Oncotarget* **8**, 40514–40532 (2017). doi: 10.18632/oncotarget.17121; pmid: 28465489
89. K. Mahadevan *et al*., RanBP2/Nup358 potentiates the translation of a subset of mRNAs encoding secretory proteins. *PLOS Biol.* **11**, e1001545 (2013). doi: 10.1371/journal.pbio.1001545; pmid: 23630457
90. C. V. Heath *et al*., Nuclear pore complex clustering and nuclear accumulation of poly(A)+ RNA associated with mutation of the Saccharomyces cerevisiae RAT2/NUP120 gene. *J. Cell Biol.* **131**, 1677–1697 (1995). doi: 10.1083/jcb.131.6.1677; pmid: 8557737
91. D. E. Neilson *et al*., Infection-triggered familial or recurrent cases of acute necrotizing encephalopathy caused by mutations in a component of the nuclear pore, RANBP2. *Am. J. Hum. Genet.* **84**, 44–51 (2009). doi: 10.1016/j.ajhg.2008.12.009; pmid: 19118815
92. K. Sell *et al*., Variable clinical course in acute necrotizing encephalopathy and identification of a novel RANBP2 mutation. *Brain Dev.* **38**, 777–780 (2016). doi: 10.1016/j.braindev.2016.02.007; pmid: 26923722
93. E. Bonnin *et al*., Biallelic mutations in nucleoporin NUP88 cause lethal fetal akinesia deformation sequence. *PLOS Genet.* **14**, e1007845 (2018). doi: 10.1371/journal.pgen.1007845; pmid: 30543681
94. S. G. Regmi *et al*., The Nuclear Pore Complex consists of two independent scaffolds. bioRxiv 2020.11.13.381947 [Preprint] (2020); doi: 10.1101/2020.11.13.381947
95. V. Zila *et al*., Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell* **184**, 1032–1046.e18 (2021). doi: 10.1016/j.cell.2021.01.025; pmid: 33571428
96. A. P. Schuller *et al*., The cellular environment shapes the nuclear pore complex architecture. *Nature* **598**, 667–671 (2021). doi: 10.1038/s41586-021-03985-3; pmid: 34646014
97. M. Hamed, B. Caspar, S. A. Port, R. H. Kehlenbach, A nuclear export sequence promotes CRM1-dependent targeting of the nucleoporin Nup214 to the nuclear pore complex. *J. Cell Sci.* **134**, jcs258095 (2021). doi: 10.1242/jcs.258095; pmid: 33589493
98. T. C. Walther *et al*., The cytoplasmic filaments of the nuclear pore complex are dispensable for selective nuclear protein import. *J. Cell Biol.* **158**, 63–77 (2002). doi: 10.1083/jcb.200202088; pmid: 12105182
99. E. Onischenko *et al*., Natively Unfolded FG Repeats Stabilize the Structure of the Nuclear Pore Complex. *Cell* **171**, 904–917.e19 (2017). doi: 10.1016/j.cell.2017.09.033; pmid: 29033133
100. L. Miorin *et al*., SARS-CoV-2 Orf6 hijacks Nup98 to block STAT nuclear import and antagonize interferon signaling. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 28344–28354 (2020). doi: 10.1073/pnas.2016650117; pmid: 33097660
101. C. J. Bley *et al*., Architecture of the cytoplasmic face of the nuclear pore, Version 1.0, CaltechDATA (2021). doi: 10.22002/D1.2209

ACKNOWLEDGMENTS

We thank the members of the Hoelz laboratory, M. Guttman, and S. Shan for insightful discussions; A. Cohen, J. Chou, R. Gulati, Y. Jeon, H. Margolis, E. Stuwe, T. Stuwe, and J. Thai for experimental support; M. Beck for sharing the ~12-Å cryo-ET reconstruction of the intact human NPC prior to publication; and V. Doye, U. Kutay, E. Hurt, and I. Mattaj for providing material. We acknowledge J. Kaiser and the scientific staff of the Stanford Synchrotron Radiation Laboratory (SSRL) Beamline 12-2 and the National Institute of General Medical Sciences and National Cancer Institute Structural Biology Facility (GM/CA) at the Advanced Photon Source (APS) for their support with x-ray diffraction measurements. **Funding:** The Molecular Observatory at the California Institute of Technology (Caltech) is supported by Donald and Judith Voet, the Gordon and Betty Moore Foundation, and the Beckman Institute. The operations at the SSRL and APS are supported by the US Department of Energy and the National Institutes of Health (NIH). GM/CA has been funded in whole or in part with federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006). S.P. and F.M.H. were supported by a PhD fellowship of the Boehringer Ingelheim Fonds. D.H.L. was supported by an NIH Research Service Award (5 T32 GM07616) and Amgen Graduate Fellowship through the Caltech-Amgen Research Collaboration. S.C., S.G.R., and M.D. were supported by National Institute of Child Health and Human Development Division of Intramural Research funding ZIAHD001902 and ZIAHD008954. A.A.K. was supported by NIH grants R01-GM117372 and P50-GM082545. A.H. was supported by a Camille-Dreyfus Teacher Scholar Award (TC-15-082) and NIH grants R01-GM117360 and R01-GM111461 and is an Investigator of the Heritage Medical Research Institute (HMRI-15-09-01) and a Faculty Scholar of the Howard Hughes Medical Institute (55108534). **Author contributions:** A.H. conceived and coordinated the study. C.J.B., S.N., G.W.M., S.P., A.T.G., X.L., S.M., S.H., F.M.H., D.H.L., B.B., A.W.T., E.J.R., A.R.C., S.C., S.G.R., T.A.S., C.A.J., M.D., A.P., A.F.P., A.A.K., and A.H. designed the research. C.J.B., S.N., G.W.M., S.P., A.T.G., X.L., S.M., S.H., F.M.H., D.H.L., B.B., A.W.T., E.J.R., A.R.C., S.C., S.G.R., T.A.S., C.A.J., and A.H. performed the research. C.J.B., S.N., G.W.M., S.P., A.T.G., X.L., S.M., S.H., F.M.H., D.H.L., B.B., A.W.T., E.J.R., A.R.C., S.C., S.G.R., T.A.S., C.A.J., M.D., A.P., A.F.P., A.A.K., and A.H. analyzed the data. S.M. and A.A.K. (synthetic antibodies) and S.C., S.G.R., and M.D. (auxin-degron cell lines) provided new reagents. A.P. and A.F.P. provided reagents and experimental guidance. C.J.B., S.N., G.W.M., S.P., A.T.G., X.L., A.P., and A.H. integrated and conceptualized the results. C.J.B., S.N., G.W.M., S.P., A.P., and A.H. wrote and revised the manuscript, with contributions from all authors. **Competing interests:** The authors declare no conflicts of interest. **Data and materials availability:** Materials generated in this study are available on request from the corresponding author. The auxin-inducible AID::NUP358 HCT116, AID::NUP358 DLD1, and NUP160::NG AID DLD1 degron cell lines are subject to a materials transfer agreement, which is available upon request. The coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with accession codes 7MNJ ($NUP358^{145\text{-}673}$), 7MNK ($NUP358^{OE}$), 7MNI ($NUP88^{NTD}$•$NUP98^{APD}$), 7MNL ($NUP358^{NTD}$•sAB-14), 7MNM ($NUP358^{NTD}$ T585M•sAB-14), 7MNN ($NUP358^{NTD}$ T653I•sAB-14), 7MNO ($NUP358^{NTD}$ I656V•sAB-14), 7MNP ($NUP358^{ZnF2}$•Ran(GDP)), 7MNQ ($NUP358^{ZnF2}$•Ran(GDP)), 7MNR ($NUP358^{ZnF3}$•Ran(GDP)), 7MNS ($NUP358^{ZnF4}$•Ran(GDP)), 7MNT ($NUP358^{ZnF5/6}$•Ran(GDP)), 7MNU ($NUP358^{ZnF7}$•Ran(GDP)), 7MNV ($NUP358^{ZnF8}$•Ran(GDP)), 7MNW ($NUP358^{RanBD\text{-}I}$•Ran(GMPPNP)), 7MNX ($Nup358^{RanBD\text{-}II}$•Ran(GMPPNP)), 7MNY ($NUP358^{RanBD\text{-}III}$•Ran(GMPPNP)), 7MNZ ($NUP358^{RanBD\text{-}IV}$•Ran(GMPPNP)), 7MO0 ($NUP50^{RanBD}$•Ran(GMPPNP)), 7MO1 ($NUP153^{ZnF1}$•Ran(GDP)), 7MO2 ($NUP153^{ZnF2}$•Ran(GDP)), 7MO3 ($NUP153^{ZnF3}$•Ran(GDP), 2.05 Å), 7MO4 ($NUP153^{ZnF3}$•Ran(GDP), 2.4 Å), and 7MO5 ($NUP153^{ZnF4}$•Ran(GDP)). PyMol and Chimera sessions containing the composite structures of the constricted and dilated human NPCs can be obtained from our website (http://ahweb.caltech.edu), and coordinates are deposited in the PDB with accession numbers 7TBL and 7TBM, respectively. Quantitative docking data, workflow code, and PyMol and Chimera sessions have been deposited in the CaltechDATA repository (*101*). **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.abm9129
Materials and Methods
Supplementary Text
Figs. S1 to S89
Tables S1 to S18
References (*102*–*153*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 22 October 2021; accepted 15 April 2022
10.1126/science.abm9129

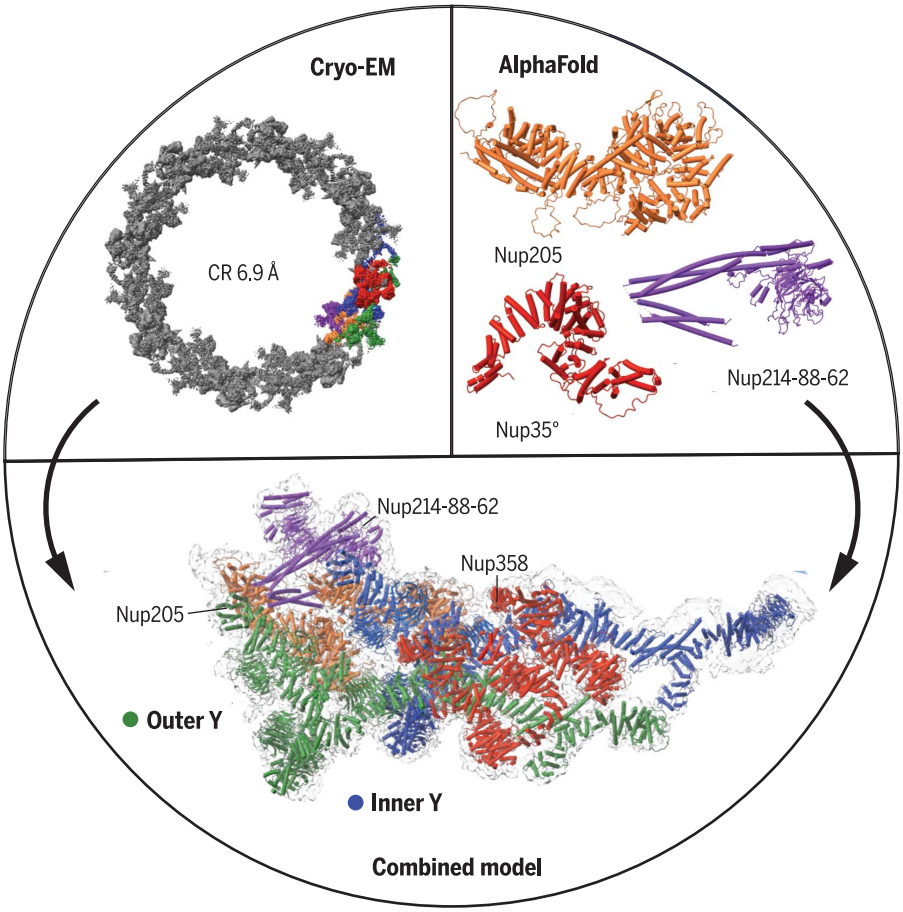
RESEARCH ARTICLE SUMMARY

NUCLEAR PORE COMPLEX

Structure of cytoplasmic ring of nuclear pore complex by integrative cryo-EM and AlphaFold

Pietro Fontana†, Ying Dong†, Xiong Pi†, Alexander B. Tong†, Corey W. Hecksel, Longfei Wang, Tian-Min Fu, Carlos Bustamante, Hao Wu*

INTRODUCTION: The nuclear pore complex (NPC) is the molecular conduit in the nuclear membrane of eukaryotic cells that regulates import and export of biomolecules between the nucleus and the cytosol, with vertebrate NPCs ~110 to 125 MDa in molecular mass and ~120 nm in diameter. NPCs are organized into four main rings: the cytoplasmic ring (CR) at the cytosolic side, the inner ring and the luminal ring on the plane of the nuclear membrane, and the nuclear ring facing the nucleus. Each ring possesses an approximate eightfold symmetry and is composed of multiple copies of different nucleoporins. NPCs have been implicated in numerous biological processes, and their dysfunctions are associated with a growing number of serious human diseases. However, despite pioneering studies from many groups over the past two decades, we still lack a full understanding of NPCs' organization, dynamics, and complexity.



Cryo-EM structure of the cytoplasmatic ring of the nuclear pore complex from *X. leavis*. The 6.9 Å map was generated with single-particle cryo-EM, and the model was built with AlphaFold structure prediction. The secondary structural elements guided EM map fitting, resulting in an almost complete model of the complex. The approach allowed the identification of five copies of Nup358 and a second copy of the trimeric Nup214-Nup88-Nup62 complex.

RATIONALE: We used the *Xenopus laevis* oocyte as a model system for the structural characterization because each oocyte possesses a large number of NPC particles that can be visualized on native nuclear membranes without the aid of detergent extraction. We used single-particle cryo–electron microscopy (cryo-EM) analysis on data collected at different stage tilt angles for three-dimensional reconstruction and structure prediction with AlphaFold for model building.

RESULTS: We reconstructed the CR map of *X. laevis* NPC at 6.9 and 6.7 Å resolutions for the full CR protomer and a core region, respectively, and predicted the structures of the individual nucleoporins using AlphaFold because no high-resolution models of *X. laevis* Nups were available. For any ambiguous subunit interactions, we also predicted complex structures, which further guided model fitting of the CR protomer. We placed the nucleoporin or complex structures into the CR density to obtain an almost full CR atomic model, composed of the inner and outer Y-complexes, two copies of Nup205, two copies of the Nup214-Nup88-Nup62 complex, one Nup155, and five copies of Nup358. In particular, we predicted the largest protein in the NPC, Nup358, as having an S-shaped globular domain, a coiled-coil domain, and a largely disordered C-terminal region containing phenylalanine-glycine (FG) repeats previously shown to form a gel-like condensate phase for selective cargo passage. Four of the Nup358 copies clamp around the inner and outer Y-complexes to stabilize the CR, and the fifth Nup358 situates in the center of the cluster of clamps. AlphaFold also predicted a homo-oligomeric, likely specifically pentameric, coiled-coil structure of Nup358 that may provide the avidity for Nup358 recruitment to the NPC and for lowering the threshold for Nup358 condensation in NPC biogenesis.

CONCLUSION: Our studies offer an example of integrative cryo-EM and structure prediction as a general approach for attaining more precise models of megadalton protein complexes from medium-resolution density maps. The more accurate and almost complete model of the CR presented here expands our understanding of the molecular interactions in the NPC and represents a substantial step forward toward the molecular architecture of a full NPC, with implications for NPC function, biogenesis, and regulation. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: wu@crystal.harvard.edu
†These authors contributed equally to this work.
Cite this article as P. Fontana *et al.*, *Science* **376**, eabm9326 (2022). DOI: 10.1126/science.abm9326

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abm9326

RESEARCH ARTICLE

NUCLEAR PORE COMPLEX

Structure of cytoplasmic ring of nuclear pore complex by integrative cryo-EM and AlphaFold

Pietro Fontana[1,2]†, Ying Dong[1,2]†, Xiong Pi[1,2]†, Alexander B. Tong[3]†, Corey W. Hecksel[4], Longfei Wang[1,2], Tian-Min Fu[1,2,5,6], Carlos Bustamante[3,7], Hao Wu[1,2]*

The nuclear pore complex (NPC) is the conduit for bidirectional cargo traffic between the cytoplasm and the nucleus. We determined a near-complete structure of the cytoplasmic ring of the NPC from *Xenopus* oocytes using single-particle cryo–electron microscopy and AlphaFold prediction. Structures of nucleoporins were predicted with AlphaFold and fit into the medium-resolution map by using the prominent secondary structural density as a guide. Certain molecular interactions were further built or confirmed by complex prediction by using AlphaFold. We identified the binding modes of five copies of Nup358, the largest NPC subunit with Phe-Gly repeats for cargo transport, and predicted it to contain a coiled-coil domain that may provide avidity to assist its role as a nucleation center for NPC formation under certain conditions.

The nuclear pore complex (NPC) regulates nucleocytoplasmic passage of biomolecules and has been implicated in numerous biological processes, with their dysfunctions associated with a growing number of diseases (*1*–*6*). An NPC is composed of multiple copies of more than 30 nucleoporins (Nups) with structural elements of stacked α-helical repeats and/or β-propellers, about a third of which also contain phenylalanine-glycine (FG) repeat sequences for selective transport of cargoes (*7*–*10*). The approximately eightfold symmetric NPC can be divided into the cytoplasmic ring (CR) at the cytosolic side, the inner ring (IR) and the luminal ring (LR) on the plane of the nuclear membrane, and the nuclear ring (NR) facing the nucleus (Fig. 1A) (*3*, *4*, *11*–*13*). Tremendous progress has been made toward unveiling the architecture of this enormous molecular machine (*11*–*20*). Here, we present the cryo–electron microscopy (cryo-EM) structure of the CR from *Xenopus laevis* oocytes.

Structure determination

We directly spread the nuclear envelopes (NEs) of actinomycin D (ActD)–treated *X. laevis* oocytes (*18*) onto Lacey grids with carbon foil on gold support and applied the Benzonase

[1]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA. [2]Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA 02115, USA. [3]Jason L. Choy Laboratory of Single-Molecule Biophysics, Institute for Quantitative Biosciences-QB3, and Chemistry Graduate Group, University of California, Berkeley, CA 94720, USA. [4]Division of CryoEM and Bioimaging, SSRL, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [5]Department of Biological Chemistry and Pharmacology, Ohio State University, Columbus, OH 43210, USA. [6]The Ohio State University Comprehensive Cancer Center, Columbus, OH 43210, USA. [7]Departments of Molecular and Cell Biology, Physics, and Chemistry, Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA.
***Corresponding author. Email: wu@crystal.harvard.edu**
†These authors contributed equally to this work.

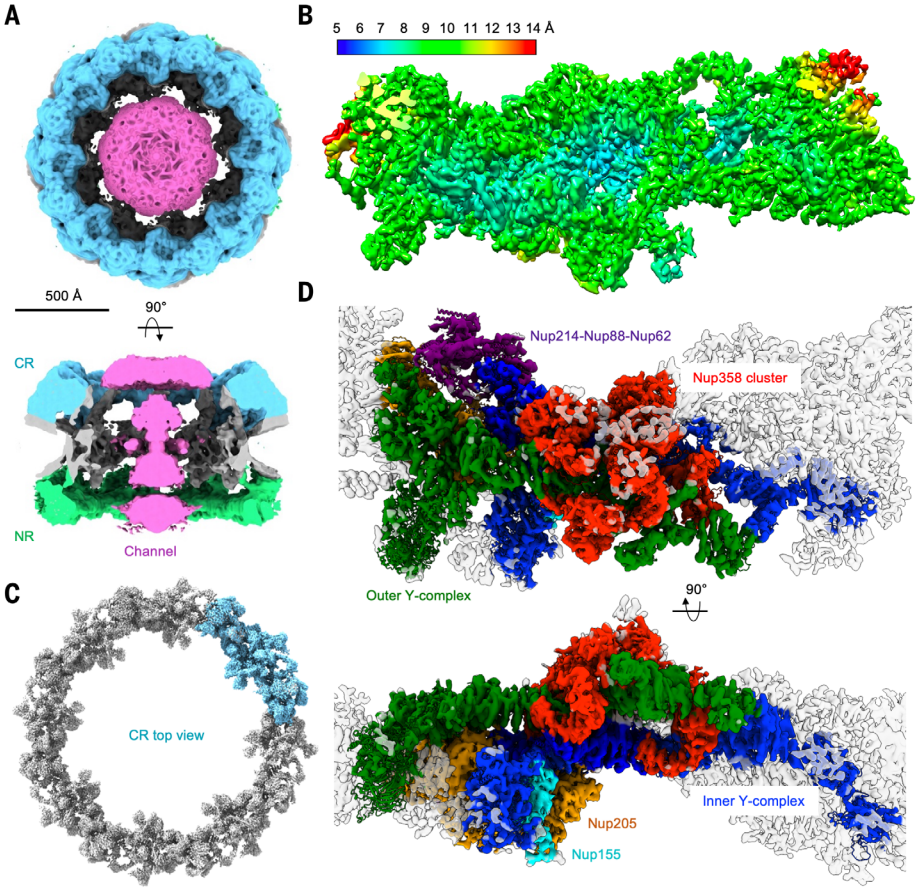


Fig. 1. Cryo-EM map of the *X. laevis* NPC. (**A**) Cryo-EM density of the *X. laevis* NPC (contour level, 3.0 σ) in top and side views, shown with CR in cyan, NR in green, IR and membrane region in gray, and the channel density in magenta. The map is eightfold symmetrized and at 19.8 Å resolution. (**B**) Cryo-EM density of a CR protomer at 6.9 Å resolution colored by local resolution. (**C**) Cryo-EM density of the *X. laevis* NPC CR ring (top view; contour level, 9.5 σ) composed from the 6.9 Å CR protomer map by assuming the eightfold symmetry. One of the CR protomers is shown in cyan. (**D**) Cryo-EM density (contour level, 4.5 σ) of a CR protomer superimposed with the final model in two orientations and colored by their model colors, with inner Y-complex in blue, outer Y-complex in green, Nup205 in orange, Nup214-Nup88-Nup62 complex in purple, Nup358 in red, and Nup155 in cyan.

nuclease to remove contaminating chromatin (fig. S1A). Cryo-EM data collection was conducted at different stage tilts and in counting mode by use of a K3 detector mounted on a Titan Krios microscope at 1.4 Å pixel size. Representative three-dimensional (3D) plots composed of the X and Y positions and the defocus levels (ΔZ) of the NPC particles in selected tilt images showed the location-dependent variation of the defocus values consistent with the tilt planes (fig. S1B). Data processing performed at the bin2 pixel size (2.8 Å) gave rise to an eightfold averaged full NPC structure, subtracted CR structure, and NR structure at 19.8, 14.6, and 14.7 Å resolutions, respectively (Fig. 1A, fig. S2, and table S1). Symmetry expansion, density subtraction, and 3D classification led to CR and NR protomers at 11.1 and 15.1 Å resolutions.

Final per-particle refinement and masking resulted in maps at 6.9 and 6.7 Å resolutions for the full CR protomer and a core region, respectively (Fig. 1, B and C; fig. S2; and table S1). The Fourier shell correlation (FSC) plots and 3D FSC plots for both maps are shown (fig. S3, A to D), as well as particle orientation distributions (fig. S3, E and F). The histograms of per-angle FSC indicated fairly isotropic resolutions along different orientations (fig. S3, C and D). The map used for density interpretation is the 6.9 Å resolution map of the full protomer. Despite the modest 6.9 Å resolution of the full CR protomer, the secondary structures, especially helices, are apparent in the maps (Fig. 1, B and C).

Model building using AlphaFold

We used the recently implemented breakthrough algorithm for protein structure prediction (AlphaFold) (*21*, *22*), mainly as the ColabFold notebook (*22*) with extended capability to predict homo- and heterocomplexes, to build a nearly complete model of the CR protomer (fig. S4), which contains the inner and outer Y-complexes, two copies of Nup205, two copies of the Nup214-Nup88-Nup62 complex, one Nup155, and five copies of Nup358 (Fig. 1D).

Because no high-resolution models of *X. laevis* Nups were available, the workflow first involved prediction of five independent models of individual Nups, which in almost all cases gave essentially the same structures (tables S2 and S3). For each prediction, we present the overall and per-residue pLDDT (predicted local distance difference test; 0 to 100, with 100 being the best), the pTM (predicted template modeling; 0 to 1, with 1 being the best), and the predicted alignment error (PAE) matrix (expected position error at residue x when the predicted and true structures are aligned on residue y, representing confidence of the relative positioning of each pair of residues or domains) (tables S2 and S3). We picked the top-ranked model by pLDDT for single proteins and by pTM for complexes in each case for density fitting unless otherwise noted. Whereas helical Nups used the prominent helical features in the maps for their fitting, Nups with mainly a β-propeller domain required prediction of binary complexes with contacting helical Nups to guide the fitting (table S4). Last, for any ambiguous subunit interactions, we predicted complex structures, which further guided model fitting of the CR protomer (table S4). *X. laevis* Nups that have a substantial region not covered by homology to structural homologs in other species include Nup107, Nup133, Nup160, Nup205, and Nup358 (tables S5 and S6 and fig. S5).

The Y-complex

The CR contains 16 copies of the Y-shaped complex (Y-complex), encircling head to tail to form the inner and outer layers of eight Y-complexes each in the ring (Fig. 1D) (*23*). Each Y-complex is composed of Nup160 and Nup37 (one short arm); Nup85, Nup43, and Seh1 (the other short arm); and Nup96, Sec13, Nup107, and Nup133 (the long stem) (Fig. 2A). Structural superposition revealed conformational differences between inner and outer Y-complexes at near Nup133 (Fig. 2B and Movie 1), likely because of the need to accommodate the different diameters at the inner and outer layers.

The AlphaFold-generated Nup160 structure fits well with the density of the inner and outer Y-complexes (Fig. 2C, fig. S5A, and tables S2 and S3). By contrast, the published homology model of *X. laevis* Nup160 [Protein Data Bank (PDB) ID 6LK8] (*14*) misses a C-terminal region (Fig. 2C), which may have led to the incorrect assignment of its density to Nup96 (Fig. 2C and fig. S5B) (*14*). Thus, building full-length models with AlphaFold may not only increase the structural accuracy of the individual subunits but also help to better assign and interpret densities.

How β-propeller Nups in the Y-complex—Nup37, Nup43, Seh1, and Sec13—fit in the CR map cannot be easily discerned. We therefore predicted structures of these Nups in complex with their contacting α-helical Nups. Seh1-Nup85, Nup43-Nup85, and Sec13-Nup96 complexes were all predicted with excellent pTM and pLDDT scores and fitted the cryo-EM density as a rigid body (Fig. 2D; fig. S5, C and D; and table S4). The Seh1-Nup85 and Sec13-Nup96 complexes exhibited hybrid β-propeller structures in which an insertion blade from the interacting helical domain completes the seven-bladed propellers (Fig. 2E and fig. S5D), as also observed in previous crystal structures of the corresponding, but partial, yeast and human complexes (*24*–*26*). AlphaFold failed to predict the Nup37-Nup160 complex (fig. S5E) (*27*), and we instead used the crystal structure to guide the Nup37 positioning in the map.

Nup205 and the Nup214-Nup88-Nup62 complex

Two AlphaFold-generated Nup205 models, which are larger than and quite different from the homologous crystal structure (*28*), were each fitted well at the channel side of the two Y-complexes to act as a bridge between them (Fig. 3A; Movie 2; fig. S6, A and B; and tables S5 and S6). The outer Nup205 runs from the C-terminal part of Nup160 to Nup85, and the inner Nup205 interacts with Nup160 at its N-terminal domain but tilts away from Nup85 at its C-terminal domain because of the interaction with the neighboring Nup214-Nup88-Nup62 complex (Fig. 3, A and B).

We fit a prominent, flag-shaped density over inner Nup85 and extending to the outer Nup85 by generating a composite model of the Nup214-Nup88-Nup62 complex (fig. S6C). The three proteins have been previously predicted to form coiled-coil interactions (*4*, *29*–*32*). According to AlphaFold, Nup88 and Nup214 also contain β-propeller domains, and complex prediction confirmed the coiled coils and agreed well with the CR map: the β-propeller of Nup88 and one end of the helical bundle as the flag base, the long helical bundle as the flagpole, and the shorter helical bundle as the banner (Fig. 3C). By contrast, the previous *X. laevis* CR structure presented only a polyalanine model for this complex (fig. S6D) (*14*). The β-propeller domain of Nup214 does not have density, likely because of a flexible linkage. A given Nup85 can only bind to either Nup205 (for outer Nup85) or the Nup214-Nup88-Nup62 complex (for inner Nup85), but not both (Fig. 3, A and D), which explains the differential modes of Nup205 interactions with the Y-complexes.

We noticed another piece of nearby density, which was previously suggested as a second Nup214-Nup88-Nup62 complex (*14*) and was fitted as such in a recent paper (*20*), which is in agreement with the expected stoichiometry from mass spectrometry data (*13*). Our density fit well with the flag base (Fig. 3D). However, the flag pole is largely missing. We do not know whether this is due to a partial disorder of this region or a lower occupancy of the second complex as a result of ActD treatment in our sample. The Nup88-Nup214-Nup62 complex resembles the *X. laevis* Nup54-Nup58-Nup62 complex anchored by Nup93 of the IR or yeast Nup49-Nup57-Nsp1 complex in its coiled-coil region (fig. S6C) (*33*, *34*), suggesting that coiled-coil structures are frequently used building blocks in NPC assembly.

The five copies of Nup358

The largest protein in the NPC, Nup358 (also known as RANBP2, or RAN binding protein 2), is composed of a largely disordered C-terminal region with FG repeats for gel-like phase formation and selective cargo passage and with binding sites for RANGAP, RAN, and other effectors (Fig. 4A) (*7*, *8*, *23*, *35*). AlphaFold predicted

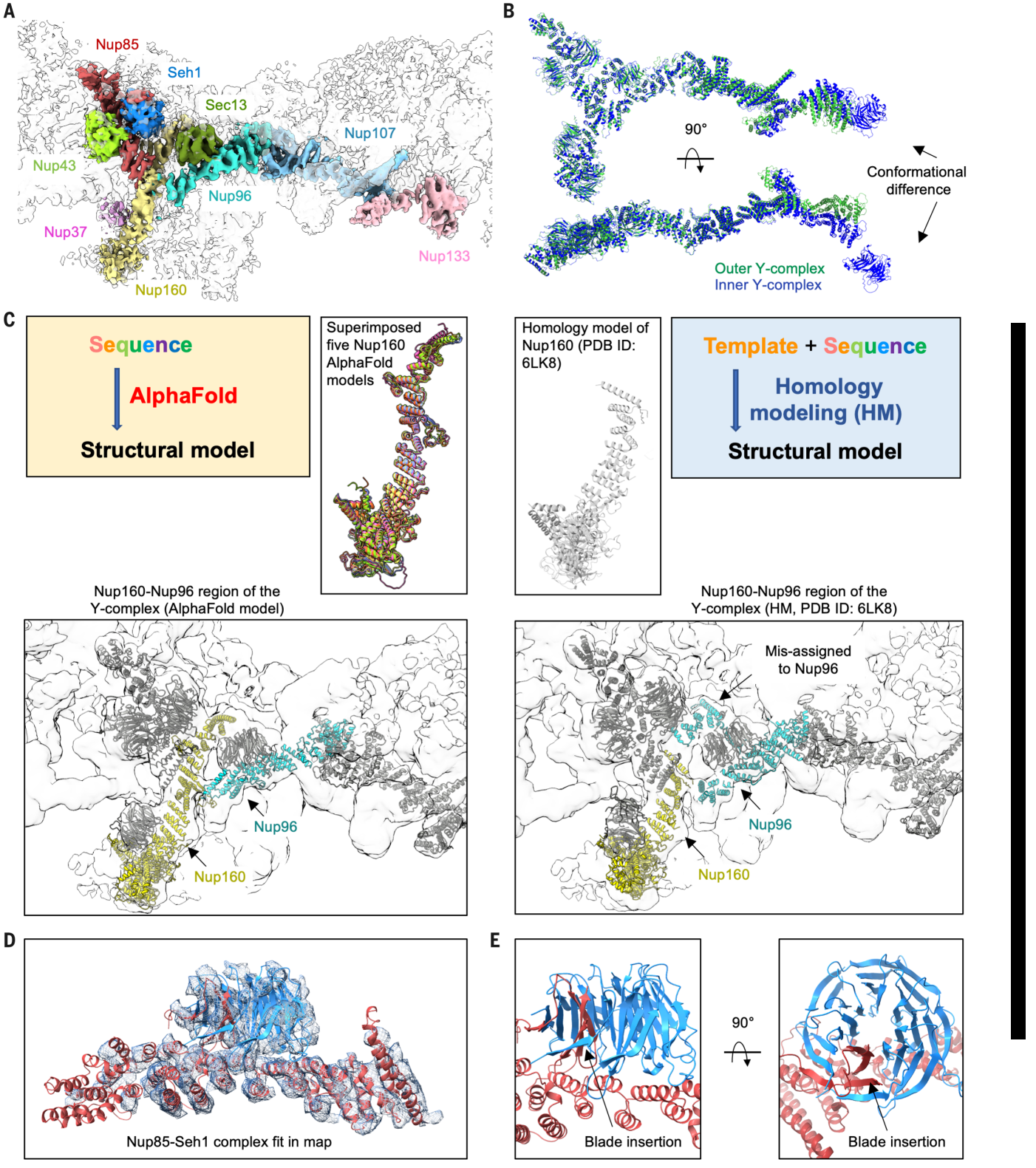
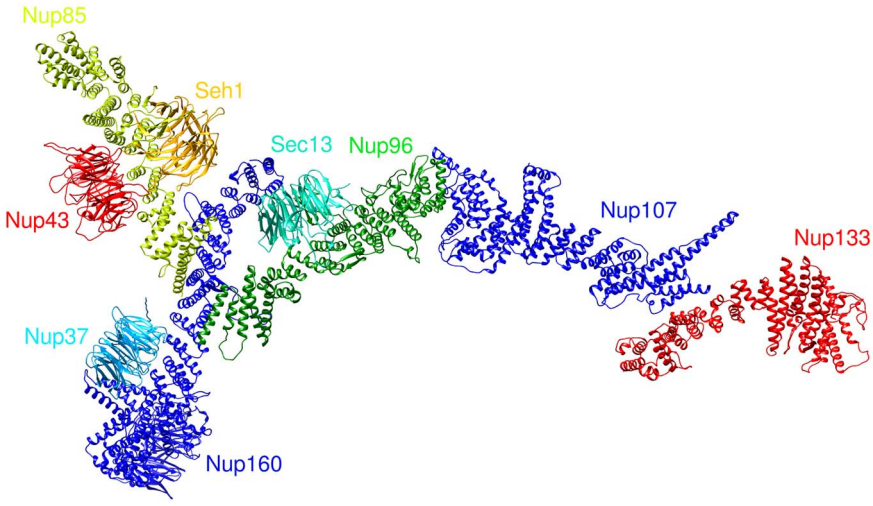


Fig. 2. Fitting of Y-complex Nups with AlphaFold. (**A**) Cryo-EM density (contour level, 8.0 σ) of the outer Y-complex colored by individual Nups. The β-propeller domain of Nup133 was not built because of lack of density. (**B**) Two views of superimposed inner Y-complex (blue) and outer Y-complex (green) by the two short arms of the Y-complexes. The distal ends of aligned Nup133 without counting the β-propeller have a distance of ~38 Å. (**C**) Comparison of AlphaFold prediction (left) and homology modeling (right) for Nup160. The cryo-EM density (contour level, 4.5 σ) and the positioning of Nup160 (yellow) and Nup96 (cyan) by the two predictions are shown at bottom. (**D** and **E**) AlphaFold-generated model of the Nup85-Seh1 complex fitted with (D) the cryo-EM density (contour level, 4.5 σ) and shown to highlight the inserted blade.



Movie 1. Conformational difference between inner and outer Y-complexes. The movie shows models of the complete Y-complexes, from 90° rotation around the horizontal axis to transition between conformations of the outer and inner Y-complexes, with the main difference at Nup133. Details are reported in Fig. 2.

the Nup358 N-terminal region as having a large α-helical domain (~800 residues), a linker, and an isolated single α-helix (Fig. 4, A and B). Previously, only the structures of a small N-terminal region (~150 residues) of human and chimpanzee NUP358 were solved (*36*) and used for homology modeling in *X. laevis* NPC (fig. S7A and tables S5 and S6) (*14*). The Nup358 globular domain is an S-shaped structure, and we identified five copies of Nup358 in the CR map (Fig. 4C and fig. S7B), which is consistent with the previous understanding of Nup358 as one of the most abundant proteins in the NPC (Fig. 4C and fig. S7B) (*4*).

The full model of Nup358 molecules shows that four of the copies clamp around the inner and outer Y-complexes near the junction of Nup96 and Nup107 (Fig. 4, D and E, and Movie 3), likely to stabilize the CR. In the outer Y-complex, clamp A contacts Nup96 and Nup107 with ~750 and 400 Å^2 buried surface area, respectively, and clamp B contacts Nup107 with ~630 Å^2 buried surface area, as calculated on the PDBePISA server (*37*). In the inner Y-complex, clamp C contacts Nup96 with only ~270 Å^2 buried surface area, and clamp D interacts with Nup107 with ~750 Å^2 buried surface area. Superposition of the inner and outer Nup96-Nup107 complexes showed that clamps B and D both contact Nup107 in a similar mode of binding, but clamps A and C are shifted significantly to account for the differences in the surface area burial (Fig. 4F). The fifth Nup358 (clamp E), situating in the center of the Nup358 cluster, contacts clamp C (~1700 Å^2) and Nup107 (~600 Å^2) of the outer Y-complex. Thus, the apparent weaker interaction to the Y-complex by clamp C is compensated by the additional interaction from clamp E.

Homo-oligomeric Nup358

We wondered whether the predicted isolated helix (Fig. 4B) following the S-shaped domain forms a coiled-coil structure, which is however invisible because of its flexible linkage. We thus used the COILS sever (*38*), which predicted up to 100% coiled-coil propensity for this helix (Fig. 5A). We then used AlphaFold to predict how the helix would assemble into oligomers. We input the number of protomers as six because coiled-coil structures with more than five subunits are very rare, and six should cover almost all possibilities. AlphaFold predicted a pentameric coiled coil plus a single helix as the top-ranked model with a pTM of 0.74 and pLDDT of 82.2. This is then followed by two trimeric coiled-coil complexes with pTMs of 0.45 and 0.44, a tetramer and a dimer with a pTM of 0.57, and last, a hexameric coiled coil with a pTM of 0.39 (Fig. 5B). The pentameric coiled coil also had the highest per-residue pLDDT scores at its core region (bluest) when displayed onto the structure (Fig. 5C).

To corroborate the AlphaFold prediction, we expressed and purified His-tagged *X. laevis* Nup358 (1 to 800, only the globular region) and Nup358 (1 to 900, with the coiled-coil region) and subjected them to gel filtration chromatography. Judging by gel filtration standards from the same column, Nup358 (1 to 800) may be consistent with a monomer, whereas Nup358 (1 to 900) may be consistent with a pentamer (Fig. 5D). A pentameric Nup358 (Fig. 5E) may help its interactions with the Y-complexes through avidity, although the potential formation of other oligomers cannot be excluded. A recent preprint reported an antiparallel tetrameric crystal structure of the coiled-coil region of human NUP358 (*39*), suggesting that Nup358 from different species may assume different modes of oligomerization.

A recurrent human mutation of NUP358, Thr^{585}→Met (T585M) (equivalent to *X. laevis* T584M), is associated with autosomal-dominant acute necrotizing encephalopathy (ADANE) (*40*, *41*). Thr^{585} is mapped to a partially buried site in direct interaction with the hydrophobic side chain of Leu^{450} (fig. S7C), suggesting that the mutation might affect the conformation of the structure and reduce its interaction with the Y-complexes. The dominant nature of this presumed loss-of-function mutation is consistent with the multimeric nature of Nup358 in which the mutant co-oligomerizes with the wild-type protein to reduce the avidity for its interaction with the Y-complexes.

Nup155 and unassigned densities

Previously, a cryo–electron tomography (cryo-ET) study of human NPC showed localization of NUP155, a linker Nup, in both the CR and the NR (*16*). The AlphaFold-predicted Nup155 structure consists of a β-propeller followed by a large helical repeat domain (Fig. 6A), in an organization similar to that of Nup160 and Nup133. The helical repeat domain fits well with the CR protomer map (Fig. 6B) and interacts with inner Nup160, burying ~750 Å^2 surface area, and with inner Nup205, burying ~310 Å^2 surface area (Fig. 6C). We wondered whether we masked out the density for the β-propeller during high-resolution refinement. The full CR map from a previous step of data processing (fig. S2) revealed density for a complete Nup155 (Fig. 6D). In this map, the β-propeller of Nup155, the neighboring inner and outer Nup160, and inner Nup133 situate inside a membrane region of the density (Fig. 6D). The β-propeller domains of Nup155 and Nup133 have been shown to possess a membrane-anchoring domain known as amphipathic lipid packing sensor (ALPS) (*42*, *43*), which consists of a short, disordered loop that may fold into an amphipathic helix on membrane (*44*).

We could not assign the identity of a piece of elongated density next to inner Nup205, Nup133, and Nup107 (fig. S8A). This density was absent from a previously deposited cryo-EM map of *X. laevis* CR (*14*) but was present in the deposited cryo-ET maps of *X. laevis* NPC treated or not with ActD (fig. S8B) (*18*). Another smaller piece of unassigned density situates adjacent to Nup358, inner Nup96, and outer Nup107 (fig. S8A). The location of this density could be explained by Nup93 as suggested by a recently released paper and a preprint (*20*, *39*). However, we were unable to properly fit Nup93 because of the weaker density.

Conclusion

Our nearly complete model of the CR of *X. laevis* NPC reveals the molecular interactions within and their biological implications. One aspect

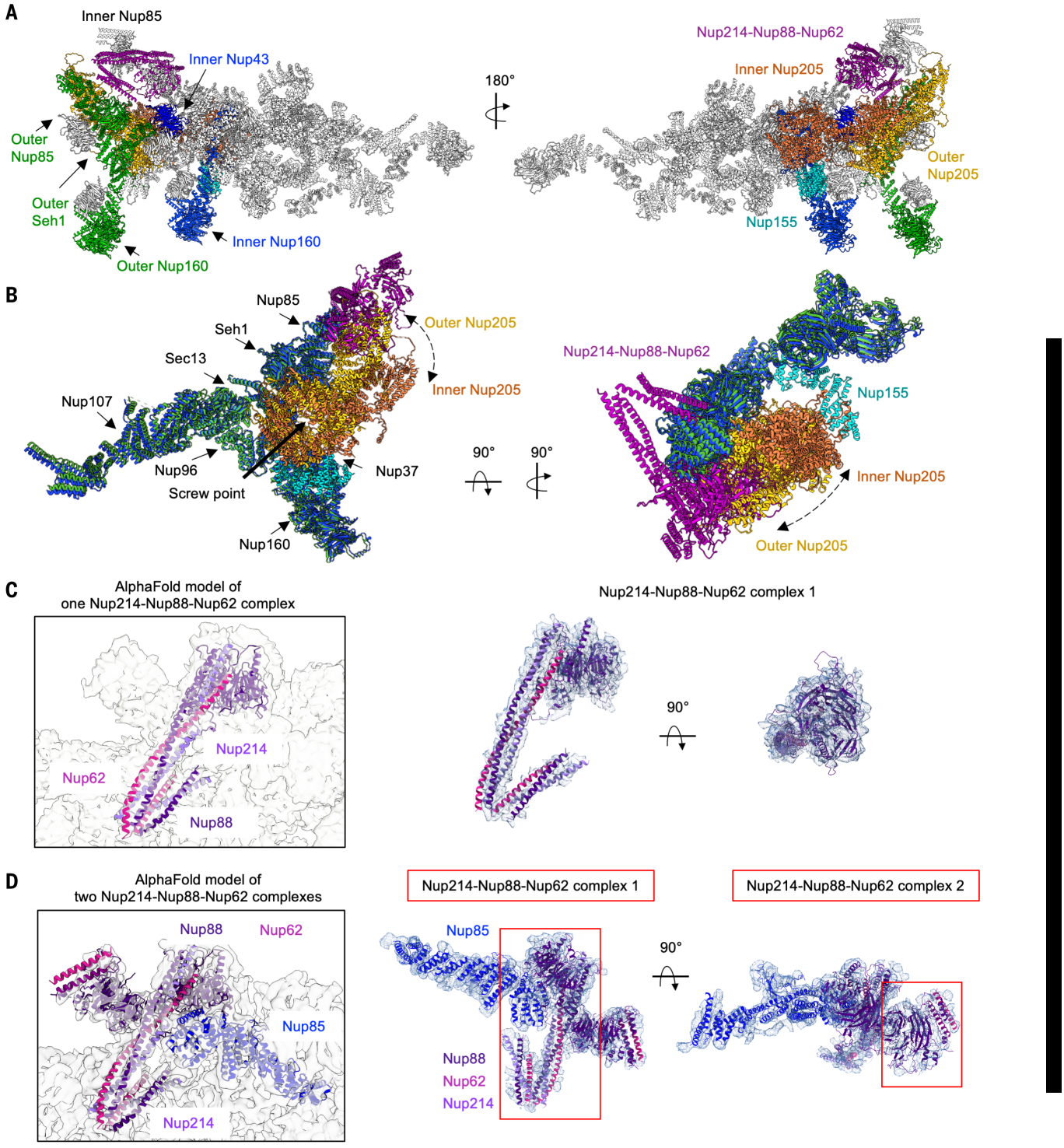
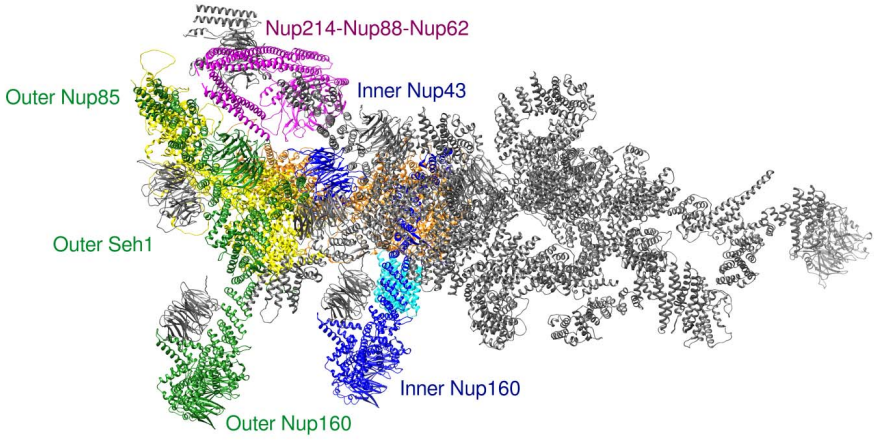


Fig. 3. Interactions mediated by Nup205 and the Nup214-Nup88-Nup62 complex. (**A**) Overall interactions of inner Nup205 (orange) and outer Nup205 (yellow) with Y-complexes. Outer Nup205 directly interacts with Nup160, Nup85, and Seh1 of the outer Y-complex and with Nup43 of the inner Y-complex. The inner Nup205 directly interacts with Nup160 of the inner Y-complex, C-terminal region of Nup155, and Nup88 β-propeller in the Nup214-Nup88-Nup62 complexes. The dashed arrows indicate the locational difference between inner and outer Nup205. (**B**) Superposition of the inner (blue) and outer (green) Y-complexes together with the bound Nup205 molecules, showing the positions of the inner (orange) and outer (yellow) Nup205 relative to Nup85 and Nup160. The N-terminal region of Nup205 binds similarly to inner and outer Nup160 molecules; the C-terminal domain binds outer Nup85 but pivots away from the inner Nup85 because of the presence of the Nup214-Nup88-Nup62 complexes. (**C**) Overview (left) and fitting (right) of the AlphaFold-predicted one Nup214-Nup88-Nup62 complex into the cryo-EM density map of NPC CR monomer (contour level, 4.5 σ). (**D**) Overview (left) and fitting (right) of the AlphaFold-predicted Nup214-Nup88-Nup62 complexes into the cryo-EM density map (contour level, 4.5 σ), with the neighboring inner Nup85. Two Nup214-Nup88-Nup62 complexes are shown.



Movie 2. Interactions formed by Nup205 and the Nup214-Nup88-Nup62 complex. The movie highlights inner and outer Nup205 and the ternary Nup214-Nup88-Nup62 complex and their interactions. The model rotates 360° along the vertical axis and 360° along the horizontal axis. Detailed interactions are reported in Fig. 3.

of the CR assembly that was unexpected is the observed asymmetry in the composition and mode of binding among Nups: the conformational differences between the two Y-complexes, the different binding modes of the two Nup205 molecules with the Y-complexes, the two Nup214-Nup88-Nup62 complexes side by side, and the five Nup358 complexes with contrasting binding modes. It will be interesting to know whether this asymmetry represents a basal state of the CR or is caused by ActD-mediated cargo deficiency, and whether it will be a common feature in the structures of the NR, IR, or LR. Our *X. laevis* NPC sample came from haploid oocytes, which may differ further from NPCs in somatic cells.

We propose that the multiple copies of Nup358 and its oligomeric coiled-coil association explain its implicated role as a key driver of NPC assembly during oogenesis in the cytosol that is different from the rapid postmitotic and the slower interphase NPC assembly (*2*). This process occurs on stacked membrane sheets of the endoplasmic reticulum (ER) termed annulate lamellae (AL), and Nup358 condensates from its FG repeats act as a fastener to spatially direct this NPC biogenesis from scratch (*2*, *45*). The additional requirement for the FG-containing Nup214 in Nup358 recruitment to the NPC (*46*) further suggests a role of condensation in NPC assembly. The oligomeric structure of Nup358 may lower the threshold for Nup358 condensation, thus helping to explain its nucleating role among the different Nups.

We also present an integrative approach to take advantage of the recent developments in cryo-EM technology (*47*, *48*) and AlphaFold structure prediction (*21*, *22*, *49*), which led to a more precise modeling of the NPC. Similar approaches were also used in the structure determination of NPCs in recently published papers or preprints (*19*, *20*, *50*–*52*). AlphaFold prediction is in contrast to structure modeling by means of homology to deposited structures that are often partial or quite dissimilar. The goal of achieving high resolution is to obtain the best model possible; incorporating information from AlphaFold in the modeling process may be analogous to what the field did previously for stereochemical restraints (*53*). With the capability for complex prediction to become more routine (*22*, *54*, *55*), we anticipate that this approach will not only assist the modeling of new structures but also help to reinterpret previous medium-resolution cryo-EM maps and become a norm in structural biology.

Materials and methods

Sample preparation for cryo-EM

X. laevis has played a key role in revealing the NPC structure because each oocyte has a large number of NPC particles (*11*, *14*, *15*, *18*, *56*). Freshly isolated stage VI oocytes of *X. laevis* in the modified Barth's saline (MBS, 10 mM HEPES at pH 7.5, 88 mM NaCl, 1 mM KCl, 0.82 mM $MgSO_4$, 0.33 mM $Ca(NO_3)_2$ and 0.41 mM $CaCl_2$) were purchased and shipped overnight from Ecocyte Bioscience US LLC. To optimize the homogeneity of the NPC sample, we incubated these oocytes with 100 μg/ml Actinomycin D (ActD) at 4°C overnight to inhibit RNA synthesis and thus RNA export for synchronization of the transport cycles (*18*). Each oocyte was poked at the animal pole using a sharp tweezer to result in the ejection of the nucleus, and transferred into a low salt buffer containing ActD (LSB, 10 mM HEPES at pH 7.5, 83 mM KCl, 17 mM NaCl and 7.5 μg/ml ActD). The nucleus was further washed to reduce the contaminating yolk in a new LSB solution. Two or three washed nuclei were then transferred to the surface of a freshly glow-discharged grid. The NE was poked open, spread using glass needles, incubated for 10 min in 10 μl of LSB supplemented with Benzonase Nuclease (Sigma Aldrich, E8263) to remove the contaminating chromatin, and subsequently washed twice with 10 μl of LSB. 3 μl LSB was added to the grid before blotting it for 3 to 5 s under 100% humidity at 4°C and plunged into liquid ethane using a Mark IV Vitrobot (ThermoFisher).

Negative staining EM

Nuclear membranes were applied on a freshly glow-discharged grid, using a Pelco EasyGlow, as described for cryo-EM sample preparation. Excess buffer was blotted on filter paper, and 6 μl of a 1% uranyl formate solution was applied for 30 s and blotted again on filter paper. Negatively stained samples were imaged on a Joel JEM1400 Transmission Electron Microscope at 120 keV.

Cryo-EM data collection

Screening and collection were performed at Stanford-SLAC Cryo-EM center (S2C2) with a Titan Krios electron microscope (Thermo Fisher Scientific) operating at 300 keV equipped with a K3 detector and a BioQuantum energy filter (Gatan, slit width 20 eV). Movies were collected in counting mode at a 1.4 Å pixel size (table S1). Because of the way the grids were made, most NPC particles would have a similar orientation with their eightfold axis perpendicular to a grid, and we were expected to use a series of stage tilt angles to alleviate this orientation bias for 3D reconstruction. Given the known knowledge that gold grids can minimize beam-induced movement (*57*), we tested a number of gold grid types with the goal of identifying one with smallest beam-induced movement that is often exaggerated at high tilt angles. These grids include Lacey carbon films on gold support, 300 mesh (Ted Pella), Quantifoil holey carbon films on gold support, R 1.2/1.3, 300 mesh (Quantifoil Micro Tools), UltrAuFoil holy gold films on gold support, R 1.2/1.3, 300 mesh (Quantifoil Micro Tools) and UltrAuFoil holy gold films on gold support overlaid with graphene (made by Wei Li Wang in the Wu lab). Lacey carbon films on gold support were shown to be the most stable and thus used for all data collection.

To alleviate the orientation bias, we initially collected datasets at stage tilts of 0°, 35°, and 45° with a total dose of 54 e/Å^2 over 40 frames for 0° and 35°, and a total dose of 79.8 e/Å^2 over 60 frames for 45°. An ideal tilt angle of 42° was then calculated using cryoEF (*58*) from a preliminary 3D reconstruction, and was used for the subsequent data collection with a total dose of 80 to 140 e/Å^2 over 80 to 120 frames. SerialEM was used for fully automated data collection, with a defocus range between −1 and −3 μm.

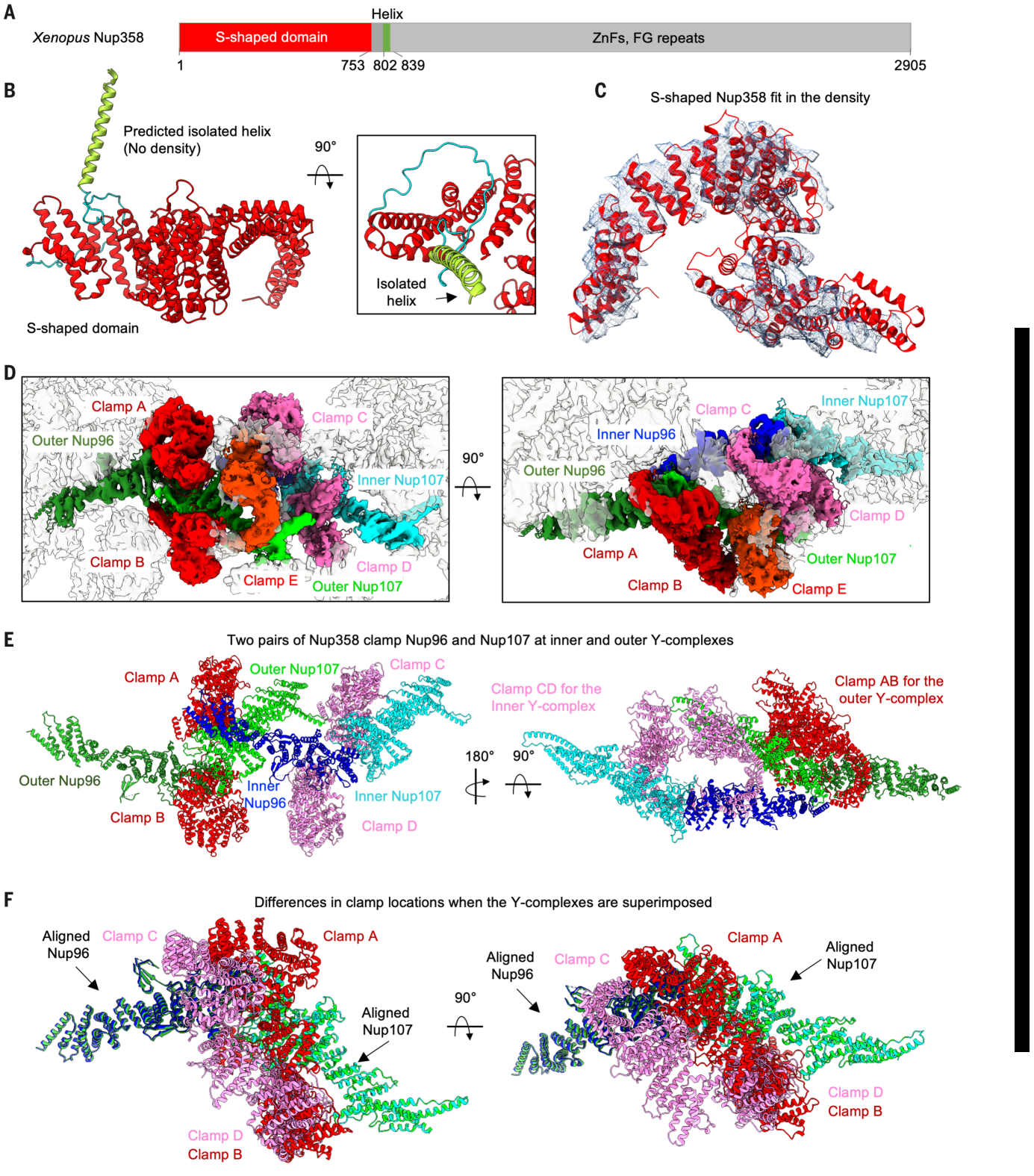
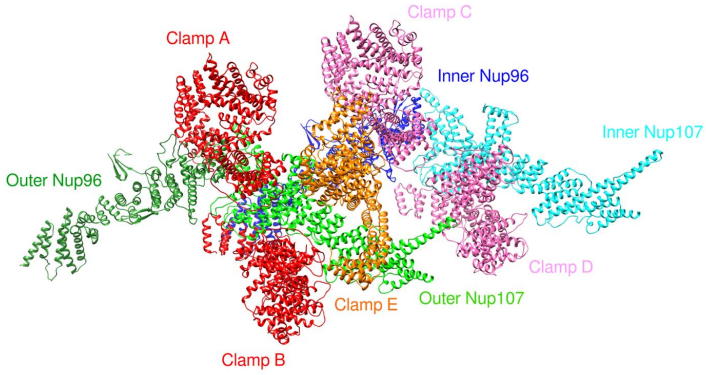
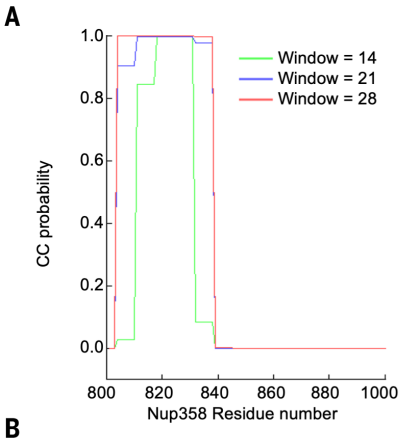


Fig. 4. Nup358 interacts with the Y-complexes as clamps. (**A**) Domain organization of *X. laevis* Nup358 and the approximate boundaries. ZnFs, zinc fingers. (**B**) AlphaFold-predicted structure of the N-terminal region of Nup358, showing the S-shaped globular domain, an isolated helix, and the flexible linker in between. (**C**) Fitting of Nup358 globular domain to the density (contour level, 8.0 σ). (**D**) The region of the map (contour level, 8.0 σ) containing five Nup358 molecules (labeled as clamps A to E) and two Y-complexes (Nup96-Nup107 complex), in two orientations. (**E**) Two Nup358 molecules each clamp around Nup96-Nup107 at the inner and outer Y-complexes. Clamps A and B (red) are for the outer Y-complex, and clamps C and D (pink) are for the inner Y-complex. The last Nup358 (clamp E, orange) contacts clamp C and Nup107 of the outer Y-complex. (**F**) Relative shifts in the clamp location on the two Y-complexes. The clamps B and D are similar in their location on Nup107, whereas clamps A and C have a shift in their position on Nup96.

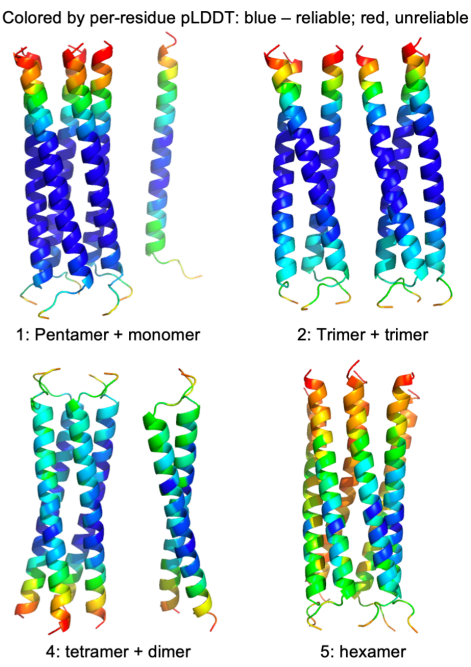


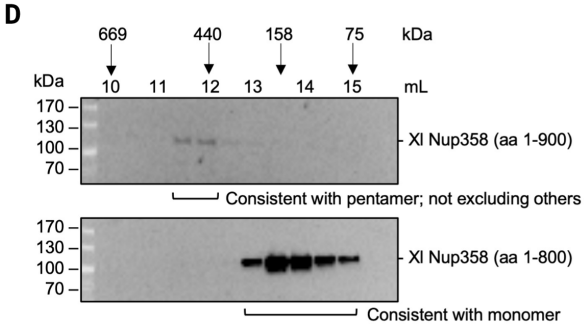
Movie 3. Interactions of Nup358 with the Y-complexes. The movie shows five copies of Nup358 and their interactions with inner and outer Nup96 and Nup107. The model zooms in to the five Nup358 clamps and then rotates 75° along the horizontal axis. Detailed interactions are reported in Fig. 4.



B

Rank	Assemblies	pTM	pLDDT
1	5 mer + 1 mer	0.74	82.2
2	3 mer + 3 mer	0.45	78.3
3	3 mer + 3 mer	0.44	77.7
4	4 mer + 2 mer	0.57	70.2
5	6 mer	0.39	54.9





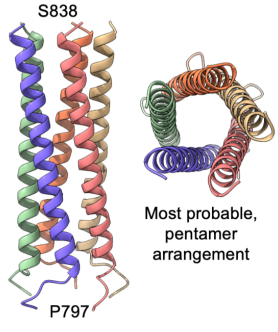


Fig. 5. Nup358 is predicted to contain an oligomeric coiled coil. (**A**) Prediction of the single helix after the S-shaped globular domain for coiled-coil propensity by using a sliding window of 14, 21, or 28 residues. (**B**) The ranked five models of six Nup358 coiled-region protomers predicted with AlphaFold and the associated pTM and average pLDDT scores. The top model contains a pentamer and a monomer, suggesting that the pentamer is the most favorable oligomer. (**C**) Ribbon diagrams of four models from (B) (ranked at 1, 2, 4, and 5) and colored by per-residue pLDDT scores. A light spectrum from blue to red corresponds to highest to lowest pLDDT scores, respectively. (**D**) Elution fractions of *X. laevis* Nup358 (1 to 900, top) and Nup358 (1 to 800, bottom) from a gel filtration column. The elution positions of several standards are shown. aa, amino acid. (**E**) The ribbon diagram of a pentamer colored by each protomer and shown in side and top views.

Cryo-EM data processing

Data processing leveraged computer support from the SBgrid Consortium (*59*). Movies were corrected by gain reference and beam-induced motion, and summed into motion-corrected and dose weighted images using the Relion 3.08 implementation of the MotionCor2 algorithm (*60*, *61*). The distribution of average motions per frame for each grid type at a given tilt angle was plotted using OriginLab (OriginPro 2017 Suite, OriginLab Corporation, Northampton, MA, USA) to evaluate grid-dependent drift performance.

The initial contrast transfer function (CTF) estimation of motion-corrected micrographs without dose-weighting was calculated by CTFFIND4 (*62*). All micrographs were manually inspected and selected based on particle uniformity and contrast, and particles were picked manually. Gctf (*63*) was then used to determine the per-particle defocus values (*63*), from which 3D plots composed of the X and Y coordinates and the CTF (Z) of the particles for selected tilt images were generated using OriginLab (OriginPro 2017 Suite, OriginLab Corporation, Northampton, MA, USA). A plane was then fit to each 3D plot of a given image (fig. S1B).

A total of 204,551 particles were manually picked, local CTF-corrected and extracted from 30,987 dose-weighted micrographs using a box size of 330 by 300 pixels at a 4× binned pixel size of 5.6 Å in RELION 3.08 (*61*). These particles were imported into cryoSPARC (*64*) to perform 2D classification, from which 124,532 good particles were selected and merged for homogeneous refinement. The published cryo-EM map of the human NPC (EMD-3103) (*16*) was low-pass filtered to 60 Å and used as the initial model. The homogeneous refinement with C8 symmetry resulted in a reconstruction at 22.1 Å. These reconstructed 124,532 particles were exported to RELION, 3.08 extracted again with a box size of 660 by 660 pixels and a binned pixel size of 2.8 Å, and imported back into cryoSPARC to re-perform 2D classification. 101,366 particles were selected for homogeneous refinement using the 22.1 Å map low-pass filtered to 40 Å as the initial model. The homogeneous refinement with C8 symmetry resulted in a 19.8 Å map. Particle density subtraction with the aligned 101,366 particles for separate processing of the CR or the NR was done in cryoSPARC. The new local refinement in cryoSPARC using the subtracted particles and a NR or a CR mask led to NR and CR maps at 14.7 and 14.6 Å resolutions, respectively.

The aligned 101,366 particles for the whole NPC were also exported to RELION 3.08 and ran auto-refine with local search and C8 symmetry, with the 19.8 map low-pass filtered to 30 Å as the initial model. The resolution of the auto-refined map was 19.5 Å. We then performed C8 symmetry expansion and density

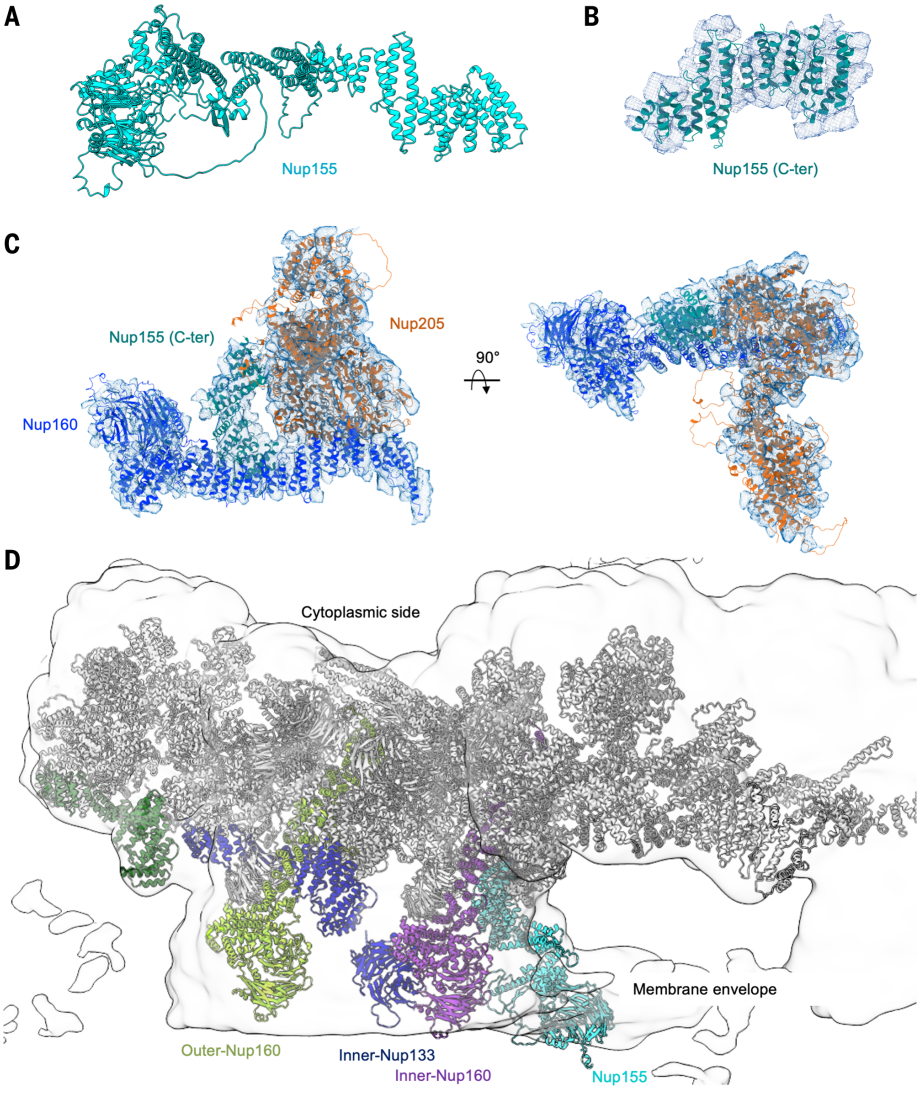


Fig. 6. Nup155 and other membrane-anchoring domains in the CR. (**A**) AlphaFold-predicted full-length Nup155. (**B**) Fitting of the C-terminal region of Nup155 into the cryo-EM density (contour level, 4.5 σ). (**C**) Interaction of Nup155 with the neighboring inner Nup160 and Nup205 (contour level, 4.5 σ). (**D**) β-propeller domains of Nup155, Nup133, and Nup160 all localize to the membrane envelope region of the cryo-EM density map of NPC CR full ring at 14.6 Å resolution (contour level, 3.0 σ).

subtraction using a CR protomer mask, and these subtracted particles were recentered and box size re-windowed to 300 by 300 pixels, all in RELION 3.1. 3D classification using a CR protomer mask, local search with 50 iterations and $K = 6$ was done on these subtracted particles. A class with 333,214 particles was selected for auto-refine with a mask and local search, reaching an 11.1 Å resolution. CTF refinement accounting beam-tilt estimation, anisotropic magnification estimation and per-particles defocus estimation and the subsequent auto-refine resulted in an improved map at 9.9 Å resolution. Additional reconstructions using a tight CR protomer mask or a tight core region mask led to maps at 8.8 and 8.4 Å resolutions. These aligned 333,214 subtracted particles were also imported into cryoSPARC to perform local CTF refinement and local refinement. The final resolutions for the CR protomer and the core region were 6.9 Å and 6.7 Å, respectively. All reported resolutions were estimated based on the gold-standard FSC = 0.143 criterion (fig. S2). All final maps were corrected and sharpened by applying a negative B factor using automated procedures in RELION 3.1. Local resolution variations of cryo-EM maps were estimated using Phenix.

Prediction of NPC subunit structures by AlphaFold

The AlphaFold structures in this study were mainly generated from the AlphaFold2 implementation in the ColabFold notebooks (*49*) running on Google Colaboratory (*21*, *22*), using the default settings with Amber relaxation (msa_method=mmseqs2, homooligomer=1, pair_mode=unpaired, max_msa=512:1024, subsample_msa=True, num_relax=5, use_turbo=True, use_ptm=True, rank_by=pLDDT, num_models=5, num_samples=1, num_ensemble=1, max_recycles=3, tol=0, is_training=False, use_templates=False). The major difference of ColabFold from the native AlphaFold2 implementation is that ColabFold uses mmseqs2 (*65*), which the ColabFold authors suggest give equivalent results (*22*). For complex prediction, sequences were entered in tandem and separated by a semicolon. For coiled coil prediction, we used homooligomer=6. Due to computing memory constraints on Google Colaboratory, we sometimes split up large proteins at disordered junctions to predict each segment separately.

AlphaFold was run once with each of the 5 trained models; the five models generated were checked for consistency, and unless specified otherwise, the top-ranked model was taken in each case for density fitting. AlphaFold computes pLDDT score and pTM score to indicate the accuracy of a prediction (*23*). We used pLDDT for ranking single protein models and pTM for ranking protein-protein complexes, as recommended by ColabFold (*22*). A predicted alignment error map between pairs of residues was also calculated for each prediction, which represents confidence in domain positioning. Confidence metrics (global and per-residue pLDDT, pTM, and PAE maps) of predictions made in this work can be found in tables S2 to S4. A few larger proteins or complexes (more than 1400 residues in total length) were run on a Boston Children's Hospital GPU cluster, by using default AlphaFold settings.

To color ribbon diagrams based on per-residue pLDDT scores (range 0 to 100, with higher being better), these scores stored at the B-factor column of the .pdb files were changed to 100-pLDDT; thus, when colored as pseudo-B-factors in Pymol (*66*), a light spectrum from blue to red corresponds to highest to lowest pLDDT scores.

Model fitting and building

Prior to beginning modeling, we used AlphaFold (*21*, *22*) to generate all models of known components of the CR using the specific *X. laevis* sequences. An initial model of the Y-complex (PDB ID: 6LK8) (*14*) was fitted into the cryo-EM density using ChimeraX (*67*), and used as a reference for manual positioning of AlphaFold-generated subunit or complex structures into the density followed by executing the "fit in map" command to refine the replacement. Flexible loops were removed to avoid steric clash. After building the two Y-complexes, we began to model the other densities. Nup205 cryo-EM density was easily recognized behind the Y-complexes due to the large size and overall

shape. Inner and outer Nup205 assume a different position due to the presence of the Nup214-Nup88-Nup62 complex in the inner Y-complex. Nup358 density was easily recognized in the presence of the generated AlphaFold model with a prominent S shape, and allowed for identification of 5 copies for each CR protomer. Nup88 density was recognized due to the β-propeller and the long α-helix. The additional density which belongs to the Nup214 β-propeller was recognized upon generation of its AlphaFold model. Building of the Nup88-Nup214-Nup62 complex was assisted by predicting the hetero-trimeric coiled coil stricture in AlphaFold, from which a composite model of the Nup88-Nup214-Nup62 complex was obtained. The final model was compared with the previous atomic model (PDB ID: 6LK8) (*14*). The model fitting quality was estimated for each subunit by the correlation coefficient in ChimeraX (*67*) and in Phenix (*68*). A value of correlation coefficient ranges from -1 to 1, with 1 as the perfect fit, and 0.5 to 1.0 as good fit. This modeling process using AlphaFold is reminiscent of the use of stereochemical information of amino acids and nucleic acids in the current practice of structural modeling (*53*) that increases model accuracy.

Nup358 expression and purification

X. laevis Nup358 constructs (residues 1-800 and 1-900) were cloned into pET21a with a C-terminal His tag. Expression was carried out in *E.coli* BL21 DE3. Briefly, cells were grown in terrific broth media, supplemented with 100 μg/ml of Ampicillin and 30 μg/ml of Chloramphenicol, until OD_{600} reached 0.6. Cells were then transferred at 4°C for 30 min before the addition of 1 mM IPTG and incubation overnight at 18°C. Cells were pelleted at 3,000 g for 20 min and resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM TCEP, 10 mM Imidazole) supplemented with a protease inhibitor cocktail. Lysis was performed by sonication and the soluble fraction was separated by centrifugation at 40,000 g for 1 hour at 4°C. The supernatant was incubated with Ni-NTA beads pre-equilibrated with lysis buffer, and purification was performed per manufacturer's recommendation. Eluted fractions were further separated by gel filtration chromatography with a Superdex 200 Increase 10/300 GL in gel filtration buffer (20 mM Hepes pH 7.4, 150 mM NaCl, 0.5 mM TCEP). Fractions were analyzed by Western blotting using an Anti-His antibody (Takara 631210). The Superdex 200 Increase 10/300 GL column was previously calibrated in gel filtration buffer using a high molecular weight kit from MW of 43 kDa to 669 kDa (Cytiva 28-4038-42).

REFERENCES AND NOTES

1. J. Fernandez-Martinez, M. P. Rout, One ring to rule them all? Structural and functional diversity in the nuclear pore complex. *Trends Biochem. Sci.* **46**, 595–607 (2021). doi: 10.1016/j.tibs.2021.01.003; pmid: 33563541
2. B. Hampoelz *et al.*, Nuclear pores assemble from nucleoporin condensates during oogenesis. *Cell* **179**, 671–686.e17 (2019). doi: 10.1016/j.cell.2019.09.022; pmid: 31626769
3. J. S. Glavy, The quest for the blueprint of the nuclear pore complex. *Protein J.* **38**, 363–376 (2019). doi: 10.1007/s10930-019-09858-z; pmid: 31410705
4. D. H. Lin, A. Hoelz, The structure of the nuclear pore complex (an update). *Annu. Rev. Biochem.* **88**, 725–783 (2019). doi: 10.1146/annurev-biochem-062917-011901; pmid: 30883195
5. A. Sali, From integrative structural biology to cell biology. *J. Biol. Chem.* **296**, 100743 (2021). doi: 10.1016/j.jbc.2021.100743; pmid: 33957123
6. P. A. Ferreira, The coming-of-age of nucleocytoplasmic transport in motor neuron disease and neurodegeneration. *Cell. Mol. Life Sci.* **76**, 2247–2273 (2019). doi: 10.1007/s00018-019-03029-0; pmid: 30742233
7. S. Frey, R. P. Richter, D. Görlich, FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science* **314**, 815–817 (2006). doi: 10.1126/science.1132516; pmid: 17082456
8. E. A. Lemke, The multiple faces of disordered nucleoporins. *J. Mol. Biol.* **428** (10 Pt A), 2011–2024 (2016). doi: 10.1016/j.jmb.2016.01.002; pmid: 26791761
9. D. Devos *et al.*, Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLOS Biol.* **2**, e380 (2004). doi: 10.1371/journal.pbio.0020380; pmid: 15523559
10. I. C. Berke, T. Boehmer, G. Blobel, T. U. Schwartz, Structural and functional analysis of Nup133 domains reveals modular building blocks of the nuclear pore complex. *J. Cell Biol.* **167**, 591–597 (2004). doi: 10.1083/jcb.200408109; pmid: 15557116
11. C. W. Akey, Interactions and structure of the nuclear pore complex revealed by cryo-electron microscopy. *J. Cell Biol.* **109**, 955–970 (1989). doi: 10.1083/jcb.109.3.955; pmid: 2768344
12. S. J. Kim *et al.*, Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475–482 (2018). doi: 10.1038/nature26003; pmid: 29539637
13. A. Ori *et al.*, Cell type-specific nuclear pores: A case in point for context-dependent stoichiometry of molecular machines. *Mol. Syst. Biol.* **9**, 648 (2013). doi: 10.1038/msb.2013.4; pmid: 23511206
14. G. Huang *et al.*, Structure of the cytoplasmic ring of the *Xenopus laevis* nuclear pore complex by cryo-electron microscopy single particle analysis. *Cell Res.* **30**, 520–531 (2020). doi: 10.1038/s41422-020-0319-4; pmid: 32376910
15. Y. Zhang *et al.*, Molecular architecture of the luminal ring of the *Xenopus laevis* nuclear pore complex. *Cell Res.* **30**, 532–540 (2020). doi: 10.1038/s41422-020-0320-y; pmid: 32367042
16. A. von Appen *et al.*, In situ structural analysis of the human nuclear pore complex. *Nature* **526**, 140–143 (2015). doi: 10.1038/nature15381; pmid: 26416747
17. K. H. Bui *et al.*, Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* **155**, 1233–1243 (2013). doi: 10.1016/j.cell.2013.10.055; pmid: 24315095
18. M. Eibauer *et al.*, Structure and gating of the nuclear pore complex. *Nat. Commun.* **6**, 7532 (2015). doi: 10.1038/ncomms8532; pmid: 26112706
19. C. W. Akey *et al.*, Comprehensive structure and functional adaptations of the yeast nuclear pore complex. *Cell* **185**, 361–378.e25 (2022). doi: 10.1016/j.cell.2021.12.015; pmid: 34982960
20. L. Tai *et al.*, 8 Å structure of the outer rings of the *Xenopus laevis* nuclear pore complex obtained by cryo-EM and AI. *Protein Cell* (2022). doi: 10.1007/s13238-021-00895-y; pmid: 35015240
21. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). doi: 10.1038/s41586-021-03819-2; pmid: 34265844
22. M. Mirdita, S. Ovchinnikov, M. Steinegger, ColabFold - Making protein folding accessible to all. bioRxiv [Preprint] (2021). https://doi.org/10.1101/2021.08.15.456425.
23. K. E. Knockenhauer, T. U. Schwartz, The nuclear pore complex as a flexible and dynamic gate. *Cell* **164**, 1162–1171 (2016). doi: 10.1016/j.cell.2016.01.034; pmid: 26967283
24. S. G. Brohawn, N. C. Leksa, E. D. Spear, K. R. Rajashankar, T. U. Schwartz, Structural evidence for common ancestry of the nuclear pore complex and vesicle coats. *Science* **322**, 1369–1373 (2008). doi: 10.1126/science.1165886; pmid: 18974315
25. E. W. Debler *et al.*, A fence-like coat for the nuclear pore membrane. *Mol. Cell* **32**, 815–826 (2008). doi: 10.1016/j.molcel.2008.12.001; pmid: 19111661
26. K. C. Hsia, P. Stavropoulos, G. Blobel, A. Hoelz, Architecture of a coat for the nuclear pore membrane. *Cell* **131**, 1313–1326 (2007). doi: 10.1016/j.cell.2007.11.038; pmid: 18160040
27. S. Bilokapic, T. U. Schwartz, Molecular basis for Nup37 and ELY5/ELYS recruitment to the nuclear pore complex. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15241–15246 (2012). doi: 10.1073/pnas.1205151109; pmid: 22955883
28. D. H. Lin *et al.*, Architecture of the symmetric core of the nuclear pore. *Science* **352**, aaf1015 (2016). doi: 10.1126/science.aaf1015; pmid: 27081075
29. F. Madeira *et al.*, The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47** (W1), W636–W641 (2019). doi: 10.1093/nar/gkz268; pmid: 30976793
30. S. M. Bailer, C. Balduf, E. Hurt, The Nsp1p carboxy-terminal domain is organized into functionally distinct coiled-coil regions required for assembly of nucleoporin subcomplexes and nucleocytoplasmic transport. *Mol. Cell. Biol.* **21**, 7944–7955 (2001). doi: 10.1128/MCB.21.23.7944-7955.2001; pmid: 11689687
31. P. Grandi *et al.*, A novel nuclear pore protein Nup82p which specifically binds to a fraction of Nsp1p. *J. Cell Biol.* **130**, 1263–1273 (1995). doi: 10.1083/jcb.130.6.1263; pmid: 7559750
32. N. Belgareh *et al.*, Functional characterization of a Nup159p-containing nuclear pore subcomplex. *Mol. Biol. Cell* **9**, 3475–3492 (1998). doi: 10.1091/mbc.9.12.3475; pmid: 9843582
33. T. Stuwe *et al.*, Architecture of the fungal nuclear pore inner ring complex. *Science* **350**, 56–64 (2015). doi: 10.1126/science.aac9176; pmid: 26316600
34. H. Chug, S. Trakhanov, B. B. Hülsmann, T. Pleiner, D. Görlich, Crystal structure of the metazoan Nup62•Nup58•Nup54 nucleoporin complex. *Science* **350**, 106–110 (2015). doi: 10.1126/science.aac7420; pmid: 26292704
35. J. Wu, M. J. Matunis, D. Kraemer, G. Blobel, E. Coutavas, Nup358, a cytoplasmically exposed nucleoporin with peptide repeats, Ran-GTP binding sites, zinc fingers, a cyclophilin A homologous domain, and a leucine-rich region. *J. Biol. Chem.* **270**, 14209–14213 (1995). doi: 10.1074/jbc.270.23.14209; pmid: 7775481
36. S. A. Kassube *et al.*, Crystal structure of the N-terminal domain of Nup358/RanBP2. *J. Mol. Biol.* **423**, 752–765 (2012). doi: 10.1016/j.jmb.2012.08.026; pmid: 22959972
37. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007). doi: 10.1016/j.jmb.2007.05.022; pmid: 17681537
38. A. Lupas, M. Van Dyke, J. Stock, Predicting coiled coils from protein sequences. *Science* **252**, 1162–1164 (1991). doi: 10.1126/science.252.5009.1162; pmid: 2031185
39. C. J. Bley *et al.*, Architecture of the cytoplasmic face of the nuclear pore. bioRxiv [Preprint] (2021). https://doi.org/10.1101/2021.10.26.465790.
40. P. Deshmukh, A. Singh, D. Khuperkar, J. Joseph, Acute necrotizing encephalopathy-linked mutations in Nup358 impair interaction of Nup358 with TNRC6/GW182 and miRNA function. *Biochem. Biophys. Res. Commun.* **559**, 230–237 (2021). doi: 10.1016/j.bbrc.2021.04.027; pmid: 33962210
41. A. Shibata, M. Kasai, A. Hoshino, T. Tanaka, M. Mizuguchi, RANBP2 mutation causing autosomal dominant acute necrotizing encephalopathy attenuates its interaction with COX11. *Neurosci. Lett.* **763**, 136173 (2021). doi: 10.1016/j.neulet.2021.136173; pmid: 34400285
42. S. J. Kim *et al.*, Integrative structure-function mapping of the nucleoporin Nup133 suggests a conserved mechanism for membrane anchoring of the nuclear pore complex. *Mol. Cell. Proteomics* **13**, 2911–2926 (2014). doi: 10.1074/mcp.M114.040915; pmid: 25139911
43. G. Drin *et al.*, A general amphipathic alpha-helical motif for sensing membrane curvature. *Nat. Struct. Mol. Biol.* **14**, 138–146 (2007). doi: 10.1038/nsmb1194; pmid: 17220896
44. S. A. Nordeen, D. L. Turman, T. U. Schwartz, Yeast Nup84-Nup133 complex structure details flexibility and reveals conservation of the membrane anchoring ALPS motif. *Nat. Commun.* **11**, 6060 (2020). doi: 10.1038/s41467-020-19885-5; pmid: 33247142
45. E. Onischenko *et al.*, Natively unfolded FG repeats stabilize the structure of the nuclear pore complex. *Cell* **171**, 904–917.e19 (2017). doi: 10.1016/j.cell.2017.09.033; pmid: 29033133
46. R. Bernad, H. van der Velde, M. Fornerod, H. Pickersgill, Nup358/RanBP2 attaches to the nuclear pore complex via association with Nup88 and Nup214/CAN and plays a supporting role in CRM1-mediated nuclear protein export.

Mol. Cell. Biol. **24**, 2373–2384 (2004). doi: 10.1128/MCB.24.6.2373-2384.2004; pmid: 14993277
47. W. Kühlbrandt, Biochemistry. The resolution revolution. *Science* **343**, 1443–1444 (2014). doi: 10.1126/science.1251652; pmid: 24675944
48. W. Chiu, M. F. Schmid, G. D. Pintilie, C. L. Lawson, Evolution of standardization and dissemination of cryo-EM structures and data jointly by the community, PDB, and EMDB. *J. Biol. Chem.* **296**, 100560 (2021). doi: 10.1016/j.jbc.2021.100560; pmid: 33744287
49. K. Tunyasuvunakool *et al*., Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021). doi: 10.1038/s41586-021-03828-1; pmid: 34293799
50. S. Mosalaganti *et al*., Artificial intelligence reveals nuclear pore complexity. bioRxiv [Preprint] (2021). https://doi.org/10.1101/2021.10.26.465776.
51. G. Huang *et al*., Cryo-EM structure of the inner ring from *Xenopus laevis* nuclear pore complex. bioRxiv [Preprint] (2021). https://doi.org/10.1101/2021.11.13.468242.
52. L. Tai *et al*., 8 Å structure of the nuclear ring of the *Xenopus laevis* nuclear pore complex solved by cryo-EM and AI. bioRxiv [Preprint] (2021). https://doi.org/10.1101/2021.11.10.468011.
53. W. A. Hendrickson, Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol.* **115**, 252–270 (1985). doi: 10.1016/0076-6879(85)15021-4; pmid: 3841182
54. R. Evans *et al*., Protein complex prediction with AlphaFold-Multimer. bioRxiv [Preprint] (2021). https://doi.org/10.1101/2021.10.04.463034.
55. I. R. Humphreys *et al*., Computed structures of core eukaryotic protein complexes. *Science* **374**, eabm4805 (2021). doi: 10.1126/science.abm4805; pmid: 34762488
56. C. W. Akey, M. Radermacher, Architecture of the *Xenopus* nuclear pore complex revealed by three-dimensional cryo-electron microscopy. *J. Cell Biol.* **122**, 1–19 (1993). doi: 10.1083/jcb.122.1.1; pmid: 8314837
57. C. J. Russo, L. A. Passmore, Electron microscopy: Ultrastable gold substrates for electron cryomicroscopy. *Science* **346**, 1377–1380 (2014). doi: 10.1126/science.1259530; pmid: 25504723
58. K. Naydenova, C. J. Russo, Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nat. Commun.* **8**, 629 (2017). doi: 10.1038/s41467-017-00782-3; pmid: 28931821
59. A. Morin *et al*., Collaboration gets the most out of software. *eLife* **2**, e01456 (2013). doi: 10.7554/eLife.01456; pmid: 24040512
60. 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
61. 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
62. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015). doi: 10.1016/j.jsb.2015.08.008; pmid: 26278980
63. 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
64. 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
65. M. Steinegger, J. Söding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017). doi: 10.1038/nbt.3988; pmid: 29035372
66. W. L. Delano, *The PyMol Molecular Graphics System* (Delano Scientific, 2002).
67. T. D. Goddard *et al*., UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018). doi: 10.1002/pro.3235; pmid: 28710774
68. P. D. Adams *et al*., PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010). doi: 10.1107/S0907444909052925; pmid: 20124702

ACKNOWLEDGMENTS

We thank W. Chiu for help with the design of data collection, M. Kirschner for initially offering to use oocytes from his laboratory, W. L. Wang for giving us graphene-coated UltrAuFoil holy gold films on gold support, A. N. Hayati and P. Sliz for running some AlphaFold predictions on Boston Children's Hospital's cluster, and H. Sharif for discussions on tilt data processing. The authors acknowledge Boston Children's Hospital's High-Performance Computing Resources BCH HPC Clusters Enkefalos 2 (E2) and Massachusetts Green High-Performance Computing (MGHPCC), which were made available for conducting the research reported in this publication. **Funding:** All cryo-EM data were collected at Stanford-SLAC Cryo-EM Center (S2C2) supported by the NIH Common Fund Transformative High Resolution Cryo-Electron Microscopy program (U24 GM129541). This work was also supported by the US Department of Energy, Office of Basic Energy Sciences, Nanomachine Program, under contract DE-AC02-05CH11231 (to C.B.); National Institutes of Health (NIH) grant R01GM032543 (to C.B.); and a postdoctoral fellowship from the Cancer Research Institute (to P.F.). **Author contributions:** Conceptualization: T.-M.F. and H.W. Cryo-EM sample preparation and optimization: P.F. and Y.D. Analysis of beam-induced motion and tilt-angle associated CTF: Y.D. and A.B.T. Data collection: P.F., C.W.H., Y.D., and X.P. Manual particle picking: P.F., Y.D., and X.P. Data processing: X.P., P.F., and Y.D. AlphaFold model generation: A.B.T., H.W., and P.F. Model fitting into density: P.F. Figure design and creation: Y.D., P.F., and X.P. Recombinant protein expression and purification: P.F. Participated in discussions: L.W. Supervision: H.W. and C.B. Writing, original draft: H.W., P.F., Y.D., X.P., and A.B.T. Writing, review and editing: H.W., P.F., Y.D., X.P., A.B.T., C.W.H., L.W., T.-M.F, and C.B. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data and materials reported in the main text and supplementary materials are available upon reasonable request. The electron density maps have been deposited in the Electron Microscopy Data Bank (EMDB) with accession numbers EMD-25817 and EMD-25818 for a CR protomer and a full CR ring built from the CR protomer map, respectively, and the atomic coordinates have been deposited in the Protein Data Bank with the accession number 7TDZ.
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.abm9326
Figs. S1 to S8
Tables S1 to S6
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 22 October 2021; accepted 3 March 2022
10.1126/science.abm9326

RESEARCH ARTICLE SUMMARY

NUCLEAR PORE COMPLEX

AI-based structure prediction empowers integrative structural analysis of human nuclear pores

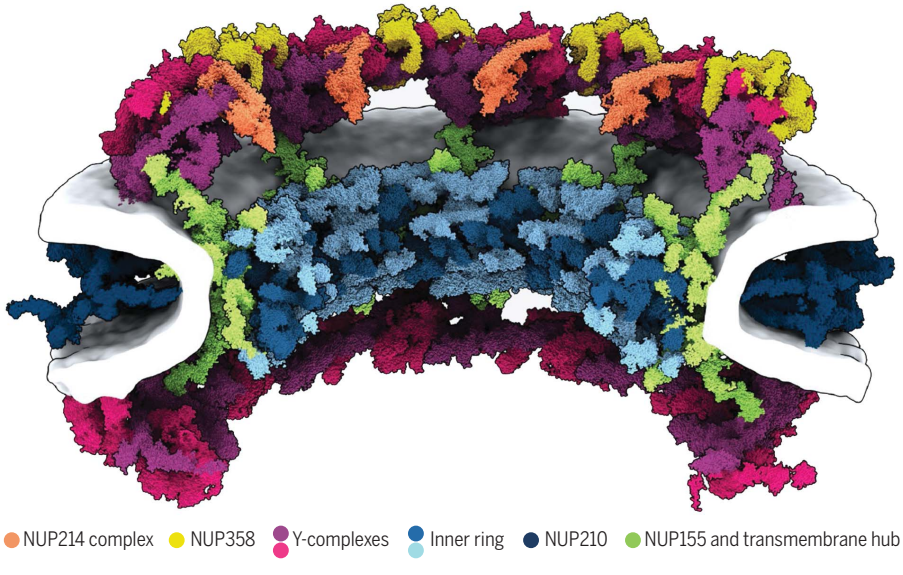
Shyamal Mosalaganti†, Agnieszka Obarska-Kosinska†, Marc Siggel†, Reiya Taniguchi, Beata Turoňová, Christian E. Zimmerli, Katarzyna Buczak, Florian H. Schmidt, Erica Margiotta, Marie-Therese Mackmull, Wim J. H. Hagen, Gerhard Hummer*, Jan Kosinski*, Martin Beck*

INTRODUCTION: The eukaryotic nucleus protects the genome and is enclosed by the two membranes of the nuclear envelope. Nuclear pore complexes (NPCs) perforate the nuclear envelope to facilitate nucleocytoplasmic transport. With a molecular weight of ~120 MDa, the human NPC is one of the largest protein complexes. Its ~1000 proteins are taken in multiple copies from a set of about 30 distinct nucleoporins (NUPs). They can be roughly categorized into two classes. Scaffold NUPs contain folded domains and form a cylindrical scaffold architecture around a central channel. Intrinsically disordered NUPs line the scaffold and extend into the central channel, where they interact with cargo complexes. The NPC architecture is highly dynamic. It responds to changes in nuclear envelope tension with conformational breathing that manifests in dilation and constriction movements. Elucidating the scaffold architecture, ultimately at atomic resolution, will be important for gaining a more precise understanding of NPC function and dynamics but imposes a substantial challenge for structural biologists.

RATIONALE: Considerable progress has been made toward this goal by a joint effort in the field. A synergistic combination of complementary approaches has turned out to be critical. In situ structural biology techniques were used to reveal the overall layout of the NPC scaffold that defines the spatial reference for molecular modeling. High-resolution structures of many NUPs were determined in vitro. Proteomic analysis and extensive biochemical work unraveled the interaction network of NUPs. Integrative modeling has been used to combine the different types of data, resulting in a rough outline of the NPC scaffold. Previous structural models of the human NPC, however, were patchy and limited in accuracy owing to several challenges: (i) Many of the high-resolution structures of individual NUPs have been solved from distantly related species and, consequently, do not comprehensively cover their human counterparts. (ii) The scaffold is interconnected by a set of intrinsically disordered linker NUPs that are not straightforwardly accessible to common structural biology techniques. (iii) The NPC scaffold intimately embraces the fused inner and outer nuclear membranes in a distinctive topology and cannot be studied in isolation. (iv) The conformational dynamics of scaffold NUPs limits the resolution achievable in structure determination.

RESULTS: In this study, we used artificial intelligence (AI)–based prediction to generate an extensive repertoire of structural models of human NUPs and their subcomplexes. The resulting models cover various domains and interfaces that so far remained structurally uncharacterized. Benchmarking against previous and unpublished x-ray and cryo–electron microscopy structures revealed unprecedented accuracy. We obtained well-resolved cryo–electron tomographic maps of both the constricted and dilated conformational states of the human NPC. Using integrative modeling, we fitted the structural models of individual NUPs into the cryo–electron microscopy maps. We explicitly included several linker NUPs and traced their trajectory through the NPC scaffold. We elucidated in great detail how membrane-associated and transmembrane NUPs are distributed across the fusion topology of both nuclear membranes. The resulting architectural model increases the structural coverage of the human NPC scaffold by about twofold. We extensively validated our model against both earlier and new experimental data. The completeness of our model has enabled microsecond-long coarse-grained molecular dynamics simulations of the NPC scaffold within an explicit membrane environment and solvent. These simulations reveal that the NPC scaffold prevents the constriction of the otherwise stable double-membrane fusion pore to small diameters in the absence of membrane tension.

CONCLUSION: Our 70-MDa atomically resolved model covers >90% of the human NPC scaffold. It captures conformational changes that occur during dilation and constriction. It also reveals the precise anchoring sites for intrinsically disordered NUPs, the identification of which is a prerequisite for a complete and dynamic model of the NPC. Our study exemplifies how AI-based structure prediction may accelerate the elucidation of subcellular architecture at atomic resolution. ■



A 70-MDa model of the human nuclear pore complex scaffold architecture. The structural model of the human NPC scaffold is shown for the constricted state as a cut-away view. High-resolution models are color coded according to nucleoporin subcomplex membership. The nuclear envelope is shown as a gray surface.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: gerhard.hummer@biophys.mpg.de (G.H.); jan.kosinski@embl.de (J.K.); martin.beck@biophys.mpg.de (M.B.)
†These authors contributed equally to this work.
Cite this article as S. Mosalaganti *et al.*, *Science* **376**, eabm9506 (2022). DOI: 10.1126/science.abm9506

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abm9506

RESEARCH ARTICLE

NUCLEAR PORE COMPLEX

AI-based structure prediction empowers integrative structural analysis of human nuclear pores

Shyamal Mosalaganti[1,2,3]†, Agnieszka Obarska-Kosinska[1,4]†, Marc Siggel[4,5,6]†, Reiya Taniguchi[1,2], Beata Turoňová[1,2], Christian E. Zimmerli[1,2], Katarzyna Buczak[2]‡, Florian H. Schmidt[2]§, Erica Margiotta[1,2], Marie-Therese Mackmull[2]¶, Wim J. H. Hagen[2], Gerhard Hummer[5,7]*, Jan Kosinski[2,4,6]*, Martin Beck[1,2]*

Nuclear pore complexes (NPCs) mediate nucleocytoplasmic transport. Their intricate 120-megadalton architecture remains incompletely understood. Here, we report a 70-megadalton model of the human NPC scaffold with explicit membrane and in multiple conformational states. We combined artificial intelligence (AI)–based structure prediction with in situ and in cellulo cryo–electron tomography and integrative modeling. We show that linker nucleoporins spatially organize the scaffold within and across subcomplexes to establish the higher-order structure. Microsecond-long molecular dynamics simulations suggest that the scaffold is not required to stabilize the inner and outer nuclear membrane fusion but rather widens the central pore. Our work exemplifies how AI-based modeling can be integrated with in situ structural biology to understand subcellular architecture across spatial organization levels.

Nuclear pore complexes (NPCs) are essential for transport between the nucleus and cytoplasm and are critical for many other cellular processes in eukaryotes (*1*–*4*). Analysis of the structure and dynamics of the NPC at high resolution has been a long-standing goal toward a better molecular understanding of NPC function. These investigations have proven challenging because of the sheer size of NPCs and their compositional and architectural complexity. With a molecular weight of ~120 MDa, NPCs form an extensive 120-nm-wide protein scaffold of three stacked rings: two outer rings—the cytoplasmic ring (CR) and the nuclear ring (NR)—and the inner ring (IR). Each ring comprises eight spokes that surround a 40- to 50-nm-wide transport channel (*5*, *6*). A single human NPC contains ~1000 copies of ~30 distinct nucleoporins (NUPs). These NUPs arrange into multiple subcomplexes, most prominently the so-called Y-complex (*7*) arranged in a head-to-tail orientation within the outer rings (*8*). The assembly of individual subcomplexes into the higher-order structure is facilitated by an as yet incompletely characterized network of short linear motifs (SLiMs) embedded into flexible NUP linkers (*9*–*13*), which have been conceived of as a molecular glue that stabilizes the scaffold. Complicating things further, the assembled scaffold is embedded into the nuclear envelope (NE). Components of the NPC scaffold interact with the NE via amphipathic helices and transmembrane domains and are believed to stabilize the fusion of the inner and the outer nuclear membranes (INM and ONM, respectively) (*14*, *15*). Finally, the FG-NUPs grafted to the scaffold form the permeability barrier filling the central channel (*16*–*18*). Their intrinsically disordered phenylalanine-glycine (FG)–rich domains challenge traditional structural biology methods.

Owing to these intricacies, the current structural models have severe shortcomings. In the case of human NPC, only 16 NUPs, accounting for ~35 MDa (30%) of the molecular weight of the complex, are included in the models (*11*, *19*, *20*). Although the repertoire of atomically resolved structures of NUPs has grown tremendously (*5*, *6*), said structures often have gaps in their sequence coverage, whereas homology models used by many studies have intrinsic inaccuracies. For some NUPs, no structures or homology models are available. Also, structural models put forward for other species are either incomplete or have limited precision (*11*, *12*, *19*, *21*–*23*). Moreover, the NPCs from many other species have a vastly reduced architectural complexity, which limits their usefulness for studying human biology (*12*, *22*–*25*). The exact grafting sites for FG-NUPs, which are crucial for understanding the transport mechanism, remain elusive. How exactly the NPC scaffold is anchored to the membrane, how it responds to mechanical cues imposed by the nuclear envelope, and if and how it contributes to shaping the membrane remain unknown. Finally, the models are static snapshots that do not take conformational dynamics into account.

In this study, we combined cryo–electron tomography (cryo-ET) analysis of the human NPC from isolated NEs and within intact cells with artificial intelligence (AI)–based structural prediction to infer a model of >90% of the human NPC scaffold at unprecedented precision and in multiple conformations. We demonstrate that AI-based models of NUPs and their subcomplexes built using AlphaFold (*26*) and RoseTTAfold (*27*) are consistent with unreleased x-ray crystallography structures, cryo–electron microscopy (cryo-EM) maps, and complementary data. We elucidate the three-dimensional (3D) trajectory of linker NUPs, the organization of membrane-binding domains, and grafting sites of most FG-NUPs in both the constricted and dilated conformations.

Results

A 70-MDa model of the human NPC scaffold

The completeness of the previous structural models of the human NPC was limited by the resolution of the available EM maps in both the constricted and the dilated states and by the lack of atomic structures for several NUPs (*19*–*21*). To improve the resolution of the constricted state of the NPC, we subjected nuclear envelopes purified from HeLa cells to cryo-ET analysis, as described previously (*19*, *21*). We collected an approximately fivefold larger dataset than what we previously published and applied a newly developed geometrically restrained classification procedure (see Materials and methods). These improvements resulted in EM maps with resolutions of 12, 12.6, and 23.2 Å, respectively, for the CR, IR, and NR (figs. S1 to S3). Next, we obtained an in cellulo cryo-ET map of dilated human NPCs in the native cellular environment within intact HeLa and human embryonic kidney 293 (HEK293) cells subjected to cryo–focused ion beam (cryo-FIB) specimen thinning (fig. S1). The dilated in cellulo NPC exhibits an IR diameter of 54 Å, compared with 42 Å in the constricted state, consistent with previous work in U2OS (*28*), HeLa (*29*), SupT1 (*30*), and, most recently, DLD-1 cells (*20*). In contrast to other species, there is no compaction along the nucleocytoplasmic axis during dilation (fig. S2) (*23*, *25*). The quality of our in cellulo map is sufficient to discern the structural features known from the constricted state, such as a double head-to-tail arrangement of Y-complexes,

[1]Department of Molecular Sociology, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany. [2]Structural and Computational Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany. [3]Life Sciences Institute and Department of Cell and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA. [4]European Molecular Biology Laboratory Hamburg, 22607 Hamburg, Germany. [5]Department of Theoretical Biophysics, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany. [6]Centre for Structural Systems Biology, 22607 Hamburg, Germany. [7]Institute of Biophysics, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany.
*Corresponding author. Email: gerhard.hummer@biophys.mpg.de (G.H.); jan.kosinski@embl.de (J.K.); martin.beck@biophys.mpg.de (M.B.)
†These authors contributed equally to this work. ‡Present address: Proteomics Core Facility, Biozentrum, University of Basel, CH-4056 Basel, Switzerland. §Present address: Institute of Science and Technology Austria, 3400 Klosterneuburg, Austria. ¶Present address: Institute of Molecular Systems Biology, Department of Biology, ETH Zurich, Zurich, Switzerland.

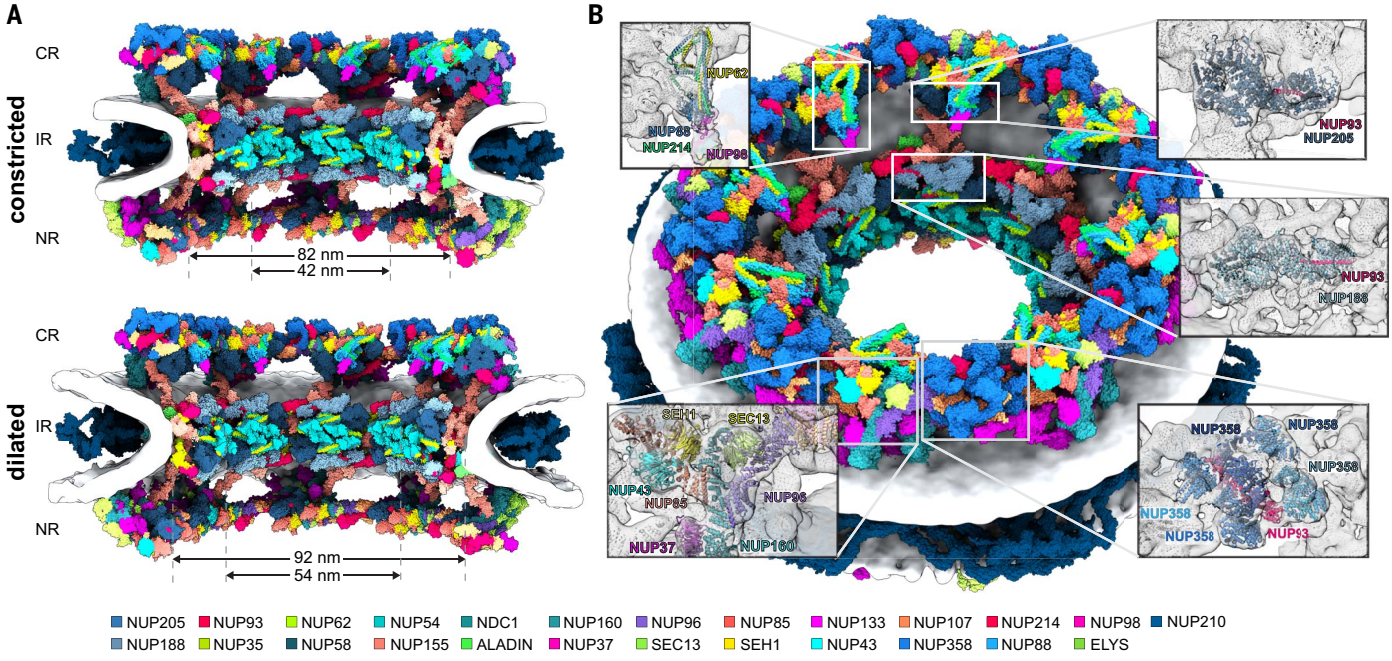


Fig. 1. Scaffold architecture of the human NPC. (**A**) The near-complete model of the human NPC scaffold is shown for the constricted and dilated states as cutaway views. High-resolution models are color coded as indicated in the color bar. The nuclear envelope is shown as a gray isosurface. (**B**) Same as (A), but shown from the cytoplasmic side for the constricted NPC. The insets show individual features of the CR and IR enlarged with secondary structures displayed as cartoons and superimposed with the isosurface-rendered cryo-ET map of the human NPC (gray).

IR subunits, and inter-ring connectors. We observe an increase in the distance between the adjacent spokes within the IR, in agreement with previous cryo-ET maps of NPCs (*20*, *22*, *23*, *25*).

To generate a comprehensive set of structural models of human NUPs, we used the recently published protein structure prediction software AlphaFold (*26*) and RoseTTAfold (*27*). We found that most of the NUPs can be modeled with high confidence scores (fig. S5 and table S1). In addition, we validated the accuracy of the models by comparison to structures from accompanying publications (*31*, *32*) of human NUP358, NUP93, NUP88, and NUP98 and of Nup205 and Nup188 from *Chaetomium thermophilum* (fig. S6 and table S1). These structures were not used as input for the modeling procedure. The AI-based models also excellently fit our EM densities, with significant *P* values and high cross-correlation scores (fig. S7, A to E). Furthermore, we used single-particle cryo-EM to determine the experimental structure of human NUP155 (fig. S8, A to C) and validated the respective AlphaFold model. Although the model and the structure do not perfectly superpose as whole chains owing to the flexibility of the protein, their local tertiary structures and side-chain conformations are highly similar [global LDDT (local distance difference test) score of 91.6] (fig. S9). Notably, the loops that were not resolved in the experimentally derived structure consistently show low predicted LDDT (pLDDT) scores (fig. S9B), further supporting the reliability of this metric.

With full-length models at hand, we could identify the positions of NUP205 and NUP188 within the scaffold, which had not been unambiguously determined in the previous human NPC (hNPC) cryo-ET maps. The AI-predicted conformation of the N-terminal domain of NUP358 fits the observed EM density better than the two x-ray structures (fig. S7H). The NUP358 localization is in agreement with previous analysis (*21*) and α-helical densities visible in the *Xenopus* EM map (*33*) (fig. S10). The full-length model of the protein ELYS, for which thus far only the N-terminal β-propeller could be placed (*21*), fits the EM map as a rigid body (fig. S7E) and confirms its binding site to each of the Y-complexes in the NR. The models of NUPs in the CR agree with the secondary structure observed in the *Xenopus* cryo-EM map (fig. S10).

The capacity of AI-based structure prediction tools to identify and model protein interfaces with high accuracy has recently been demonstrated (*34*–*36*). We therefore attempted to model NUP interfaces using the ColabFold software, a version of AlphaFold adopted for modeling protein complexes (*35*). We found that ColabFold predicted several NUP subcomplexes with interdomain confidence scores that correlated with the accuracy of the models, while negative controls with nonspecific interactions yielded low confidence models (figs. S11 to S14). The models of these subcomplexes not only reproduced their respective, already available x-ray structures but also agreed with newly resolved x-ray structures (*31*, *32*) (tables S1 and S2) and exhibited physical parameters similar to real interfaces (table S3). Specifically, x-ray structures of *C. thermophilum* Nup205 and Nup188 in complex with Nup93 as well as Nup93 in complex with Nup35 are consistent with the human ColabFold model (fig. S6 and table S2). These structures represent proteins in complex with the respective SLiMs and form relatively small interfaces. However, for larger subcomplexes we also obtained structural models that convincingly fit our cryo-ET maps (fig. S14). For example, the structure predicted for the so-called central hub of the Y-complex was consistent with the organization seen in fungal x-ray structures and explained additional density within the cryo-ET map specific to the human NPC. Our model of the Y-complex hub includes a previously unknown interaction between NUP96 and NUP160 (fig. S11). ColabFold built a model of the NUP62 complex that has high structural similarity to the fungal homolog (table S2) and fits the EM map with significant *P* values (fig. S14), even though no structural templates were used for modeling. We were also able to obtain a trimeric model of the small arm of the Y-complex comprising NUP85, SEH1, and NUP43. The model fits the EM map with significant *P* values, confirming the known structure of NUP85-SEH1 interaction (table S2) and revealing how NUP43 interacts with NUP85

(fig. S14). In the case of the NUP214 complex, for which no structures are available, the ColabFold model is highly consistent with the rather distinctively shaped EM density (fig. S14). The interface between NUP214 and NUP88 that is biochemically validated in (*31*) has also been predicted with high structural similarity to the equivalent interface between homologs from *C. thermophilum* and budding yeast (table S2).

With the cryo-EM maps and the repertoire of structural models of individual NUPs and their subcomplexes, we built a nearly complete model of the human NPC scaffold (Fig. 1A; supplementary text in the supplementary materials; and Materials and methods). The individual components are detailed in table S4. We used the previous model (*19*, *21*) as a reference for modeling the scaffold of the constricted state and replaced all previously fitted domains with human AlphaFold and ColabFold models. We then added the remaining newly modeled subunits by systematic fitting to the EM map and refinement using Assembline (*37*) (figs. S14 and S15). In addition to fitting the models, we added several disordered linkers that connect spatially separated domains and SLiMs within the NPC (figs. S16 to S18). We then built the model of the dilated state by fitting the constricted NPC model into the dilated NPC map and refining the fits using Assembline. The resulting models (Fig. 1) include 25 of the ~30 human NUPs (fig. S19 and table S4). The protein regions explicitly included in the models account for 70 MDa of the molecular weight of the NPC (>90% of the scaffold molecular weight), compared with 16 NUPs and 35 MDa (46% of the scaffold weight) of the previous model, and largely account for the EM density observed in the constricted and dilated states.

This model yields new insights into the organization of the human NPC (Fig. 1). Within the IR, NUP188 and NUP205 localize to the outer and inner subcomplexes, respectively, consistent with previous analysis in other species (*12*, *22*, *23*). Furthermore, we localized two copies of NUP205 in the CR and one in the NR (*33*), thus resolving previous ambiguities (*11*, *19*, *21*). Two previously undetected copies of NUP93 bridge the inner and outer Y-complexes in both the CR and NR, with an inherent C2 symmetry. This observation is consistent with biochemical experiments that initially identified interactors of NUP93 in the outer rings (*38*). The copy of NUP93 in the CR is located underneath the NUP358 complex, further corroborating a role of NUP358 in stabilizing the higher-order structure (*21*). Yet another copy of NUP93 that is specific to the CR bridges the inner Y-complexes from two consecutive spokes. This is consistent with an additional copy of NUP205 in the CR as compared with the NR, because NUP93 and NUP205 heterodimerize through a SLiM within the extended N terminus of NUP93 (see next section; fig. S10) (*31*, *32*). The AI-based model of the NUP214 subcomplex interacts with NUP85, which points toward the central channel, likely to optimally position the associated helicase that is crucial for mRNA export.

Linker NUPs fulfill dedicated roles of spatial organization within the higher-order assembly

Because the exact 3D trajectory of the linkers through the NPC scaffold was unknown, it remained difficult to understand their precise structural role beyond conceptualization as molecular glue. In our model, AI-based models of human NUP-SLiM subcomplexes allowed us to map the anchor points of the linkers to the scaffold. The AI-based models correctly recapitulated SLiM interactions known from x-ray structures but also revealed previously unknown human NUP-SLiM interactions. In comparison to the x-ray structures, the AI-based models more extensively covered the structured domains, thus reducing the length of the linkers and restricting their possible conformational freedom within our model.

To generate a connectivity map of the NUP linkers (Fig. 2), we used a multistep procedure. First, we calculated all geometrically possible connections. Next, we eliminated linker combinations that were too distant, caused steric clashes, or were combinatorially impossible. Finally, we used Assembline to model the remaining linkers in explicit atomic representation for both constricted and dilated states (Fig. 2, figs. S16 to S18, supplementary text, and Materials and methods).

The resulting connectivity map (Fig. 2A) reveals that the NUP35 linker regions bridge neighboring spokes of the IR. In our model, the NUP35 dimer is positioned into previously unassigned EM density between spokes (fig. S14), and each of the two copies reaches out with its SLiMs to NUP155 and NUP93 of the adjacent spokes (Fig. 2A and fig. S16). The NUP35 dimer, which is critical during early NPC biogenesis (*39*), thus functions as an architectural organizer for the IR membrane coat in a horizontal direction along the membrane plane.

In contrast, the connectivity map demonstrates that the linkers at the N terminus of the NUP93 copies that connect anchor points at NUP205 or NUP188, and NUP62 complex in the IR, cannot reach across spokes and thus connect subunits within a single IR subcomplex inside of the same spoke (fig. S17). Thereby, the two outer copies bind to NUP188, while the two inner copies bind to NUP205. Thus, NUP93 acts as an architectural organizer within, but not across, spokes.

In the CR and NR, the linkage between NUP93 and NUP205 is geometrically possible (fig. S18). This linkage suggests the similar architectural design of the respective complexes in which the NUP93 SLiM that binds the NUP62 complex could also facilitate linkage to the homologous NUP214 complex, although the corresponding structural information is still missing. The duplication of NUP205 and NUP93 in the CR is suggestive of yet another copy of the NUP214 complex that is not well resolved in the cryo-EM map of the constricted state and thus remains to be further investigated. This analysis is consistent with the biochemical analysis and structural modeling in the accompanying publication (*32*). In conclusion, the individual linker NUPs specialize in dedicated spatial organization functions responsible for distinct aspects of assembly and maintenance of the NPC scaffold architecture.

A transmembrane interaction hub organizes the interface between outer and inner rings

Several types of structural motifs associate the NPC scaffold with the membrane. The spatial distribution of the amphipathic helices and membrane-binding loops harbored by NUP160, NUP155, and NUP133 across the scaffold has been previously revealed (*19*). The analysis of the protein linkers in our model allowed the mapping of an approximate location of the amphipathic helices of NUP35. In addition to these motifs, the human NPC contains three transmembrane NUPs, the precise location of which remains unknown.

Among the three human transmembrane proteins, NDC1 is the only one that is conserved across eukaryotes (*40*). NDC1 is known to interact with the poorly characterized scaffold NUP ALADIN (*41*, *42*). We confirmed this interaction using proximity labeling mass spectrometry with BirA-tagged ALADIN and identified NDC1 and NUP35 as the most prominently enriched interactors (fig. S20). NDC1 is predicted to comprise six transmembrane helices followed by a cytosolic domain containing mainly α helices, whereas ALADIN is predicted to have a β-propeller fold. Structures of both NDC1 and ALADIN, however, remain unknown. Using AlphaFold/Colabfold, we could model the structures both as monomers and a heterodimeric complex with high-confidence scores (figs. S5 and S11). Systematic fitting of the heterodimeric models to the EM map unambiguously identified two locations within the IR (fig. S14). The EM density was not used as a restraint for modeling, but it matches the structure of the model and is consistent with the only patches of density spanning the bilayer (Fig. 3, fig. S14, and supplementary text), therefore further validating the model. The two locations are C2-symmetrically equivalent across the nuclear envelope plane, thus assigning two copies of ALADIN and NDC1 per spoke, corroborating experimentally determined stoichiometry (*43*). The identified locations are close to the

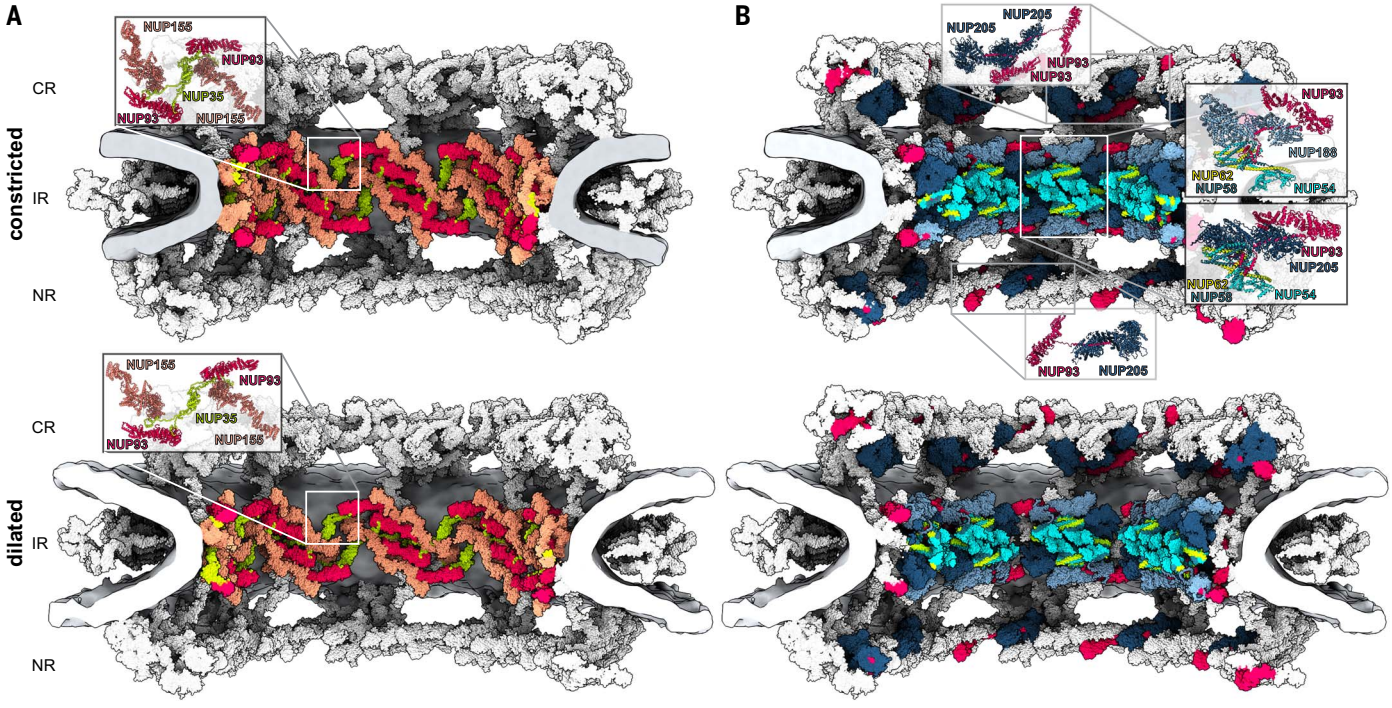


Fig. 2. The connectivity of protein linkers within the human NPC. (**A**) The NUP35 dimer interconnects adjacent spokes across different subcomplex species, thus facilitating cylindrical assembly of the IR in both constricted (top) and dilated (bottom) states. The NUP205, NUP188, NUP62 complex, and the N terminus (amino acids 1 to 170) of NUP93 are hidden from view to expose NUP35. (**B**) The N terminus of NUP93 isostoichiometrically connects the subunits NUP205, NUP188, NUP62, NUP58, and NUP54 within the same subcomplex species. Insets show NUP93 connectivity, highlighting its interaction with two copies of NUP205 in the CR, two copies each with NUP205 and NUP188 in the IR, and a single copy of NUP205 in the NR. The respective subunits are color coded as in Fig. 1, while all other subunits and the nuclear membranes are shown in gray.

membrane-binding N-terminal domains of NUP155 and amphipathic helices of NUP35, which is also consistent with our proximity labeling data (fig. S11) and previous functional analysis (*44*). The proximity labeling data prominently identified Gle1, which was previously shown to interact with NUP155 (*45*). Using AlphaFold, we predicted an interaction between the C terminus of NUP155 and the N-terminal SLiM in Gle1 with high confidence scores (fig. S20B). We also predicted an ALADIN-binding SLiM in NUP35, located between the amphipathic helix and NUP155-binding SLiM (Fig. 3 and fig. S11). We therefore propose that, together with NUP155, NUP35, ALADIN, and NDC1 form a transmembrane interaction hub that anchors the inner membrane coat of the IR and orients the NUP155 connectors toward the outer rings. The central position of ALADIN within the NPC might explain functional consequences of mutations in ALADIN that are implicated in triple A syndrome (*46*–*48*) and is consistent with the absence of ALADIN in fungi, which lack the NUP155 connectors (*24*).

We next examined the structure of NUP210, which contains a single-pass transmembrane helix and is the only NUP that primarily resides in the NE lumen. This NUP is composed of multiple immunoglobin-like domains and is thought to form a ring around the NPC within the NE lumen (*49*), but the structure of the ring has only been modeled in fungi. We used RoseTTAfold to model full-length NUP210 and obtained an elongated model with clearly defined interfaces between consecutive domains. This model fitted the density sufficiently well to allow tracing NUP210 monomers in the cryo-EM map of the *Xenopus laevis* NPC, which is superbly resolved in the luminal region (*50*). Thus, we could assign eight copies of NUP210 per spoke (fig. S21). Modeling of individual NUP210 fragments and inter-NUP210 interactions with AlphaFold/ColabFold (fig. S21 and table S1) led to a composite model that explained the entire density of the luminal ring of the human and *Xenopus* cryo-EM maps, including the C-terminal transmembrane helix. The helix is long enough to span the NE and reaches the IR in the vicinity of the NDC1/ALADIN/NUP155/NUP35 transmembrane interaction hub. This location is consistent with known interactions of NUP210 homologs (*51*) and our proximity labeling data (fig. S20). The model of the NUP210 ring also matches the luminal density visible in our in cellulo EM map, allowing us to model the ring in the context of both constricted and dilated NPC (Fig. 1 and fig. S21). To further confirm the NUP210 assignment in human cells, we deleted *NUP210* in HEK293 cells using CRISPR-Cas9 and analyzed the structure of the NPCs in cellulo using cryo-ET. The resulting map indeed showed a lack of the luminal ring density (fig. S4). The NPC scaffold in the resulting map appears unchanged overall, including its diameter, suggesting the NUP210 is not required for faithful NPC assembly.

Our model includes all known membrane-binding domains except for the cell type–specifically expressed POM121—the precise location of which remains unknown within the NPC, and neither AlphaFold nor RoseTTAfold could build structural models with high confidence. The resulting membrane association map reveals that the membrane-binding β-propellers of the Y-complex (NUP160 and NUP133) and the IR (NUP155) are distributed as multiple pairs over the entire scaffold, whereby they follow a well-defined pattern. They form an overall Z-shaped outline within an individual spoke (Fig. 3). The NDC1/ALADIN/NUP155/NUP35 membrane-binding hub is situated at the interface of the IR with the outer rings and is distinct from the additional NUP155 pair at the NE symmetry plane. The membrane-binding motifs arrange in similar clusters in both the constricted and dilated state. Their relative arrangement does not change uniformly

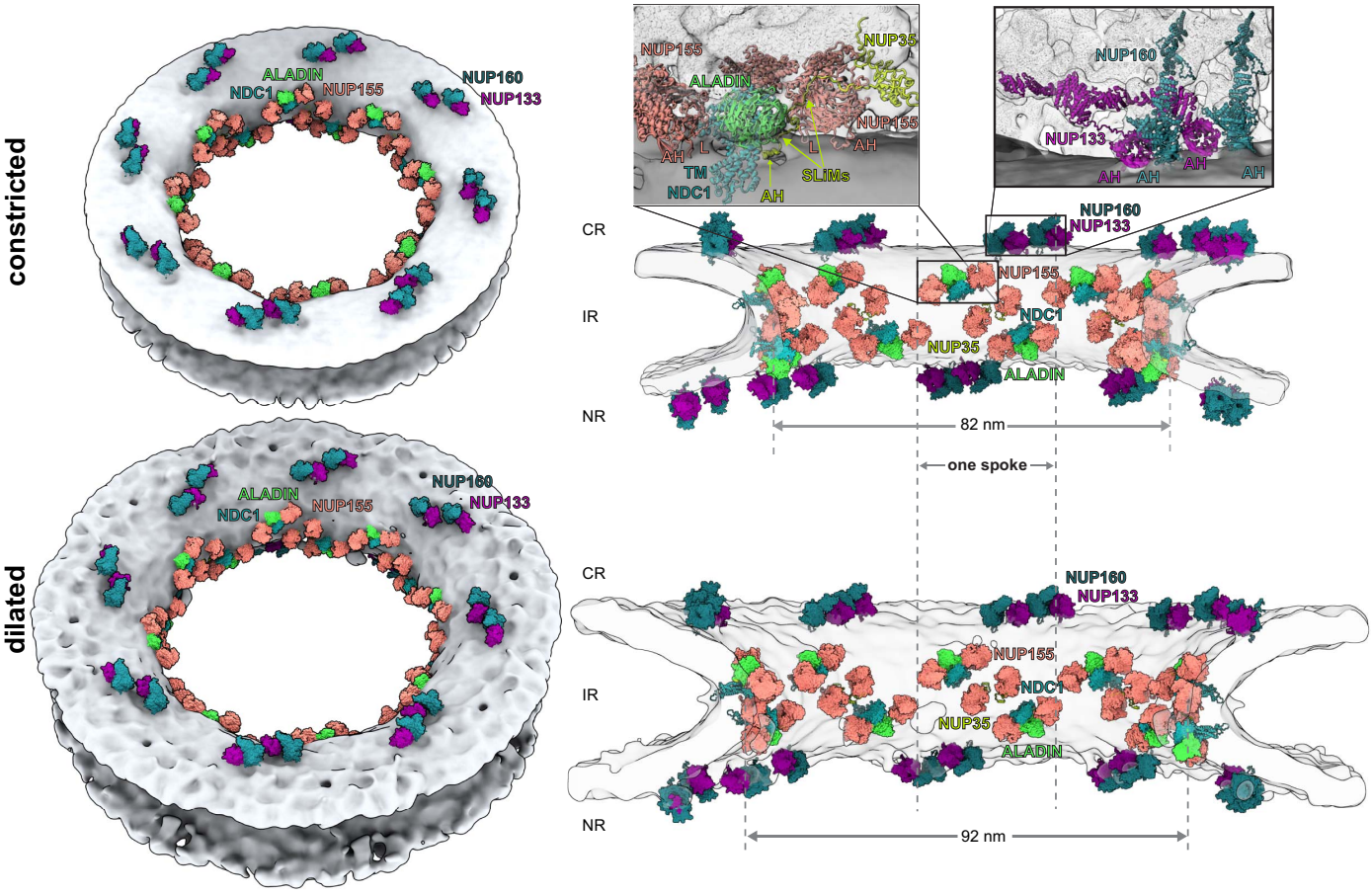


Fig. 3. The membrane-anchoring motifs of the human NPC are distributed over the entire scaffold. The membrane-binding β-propellers of the Y-complex and IR complex are shown color coded and arranged as pairs of the respective inner and outer copies. ALADIN and NDC1 form a transmembrane interaction hub with the inner and connector copies of NUP155, which is shown enlarged in the inset in the cut-away side view (right). The nuclear membranes are shown as a gray isosurface. AH, amphipathic α helix; L, loop; TM, transmembrane; SLiMs, short linear motifs.

during dilation, rather the relative distances within the spokes remain constant while the spacing of the spokes increases (Fig. 3).

The NPC scaffold prevents membrane constriction in the absence of membrane tension

It has been a long-standing view that the scaffold architecture of the NPC has evolved distinct membrane-binding motifs to stabilize the membrane at the fusion of the INM and ONM (*14*, *52*). To test the contribution of the scaffold architecture to membrane curvature, we used molecular dynamics (MD) simulations using a coarse-grained Martini force field (*53*, *54*). We first simulated a double-membrane pore without proteins with an initial pore diameter and membrane spacing as seen in the constricted NPC cryo-ET map. We found that the pore constricts during 1-μs simulations and stabilizes once the radii of the INM or ONM and the NE hole are the same, such that the mean curvature nearly vanishes (Fig. 4A, fig. S23, and Movie 1). This is in line with Helfrich membrane elastic theory, which predicts a catenoid-like pore shape with equal radii of curvature at the pore center as the lowest energy structure, and an energetic cost of ~500 kJ/mol to widen the pore (supplementary text). Notably, the opening of the relaxed double-membrane pore is considerably smaller than even the most constricted NPC conformation. The NPC scaffold thus keeps the pore wider than it would be without the scaffold.

These findings would predict that the nuclear membranes push against the NPC scaffold even in the most constricted state, which is in agreement with experimental data (*23*). To examine the effects of this tension on the NPC, we generated an NPC scaffold model with explicit membrane and water as a solvent and ran 1-μs MD simulations (Fig. 4B, figs. S23 to S27, and Movies 2 and 3). We omitted the luminal NUP210 and focused our analysis on the architecturally important NR, CR, and IR. In these simulations, we found that the membrane pore wrapped tightly around the IR plane, adopting an octagonal shape (Fig. 4C). Similarly tight wrapping and octagonal shapes have been seen in the previous EM analyses of NPCs (*21*, *55*, *56*). We also observed that the diameter of the NPC scaffold constricted by ~9% (Fig. 4B). We attribute this tightening primarily to mechanical tension in the pore widened beyond the catenoid shape. First, we observed similar contraction in simulations with rescaled protein-protein interactions (fig. S27). Second, by applying lateral tension on the double membrane, we could maintain the pore width or widen it (Fig. 4B and fig. S23B). At even higher tension, the membrane spontaneously detached from the NPC scaffold (fig. S28). Taken together, our data support a model in which the role of the NPC scaffold is not to stabilize the membrane fusion per se but rather to widen the diameter of the membrane hole without necessitating a wider envelope.

Discussion

We have built a 70-MDa model of the human NPC scaffold in the constricted state (smaller diameter) as adopted in purified nuclear envelopes and in the dilated state as adopted in cells, whereby recent work in fungi has identified constricted NPCs inside of cells under

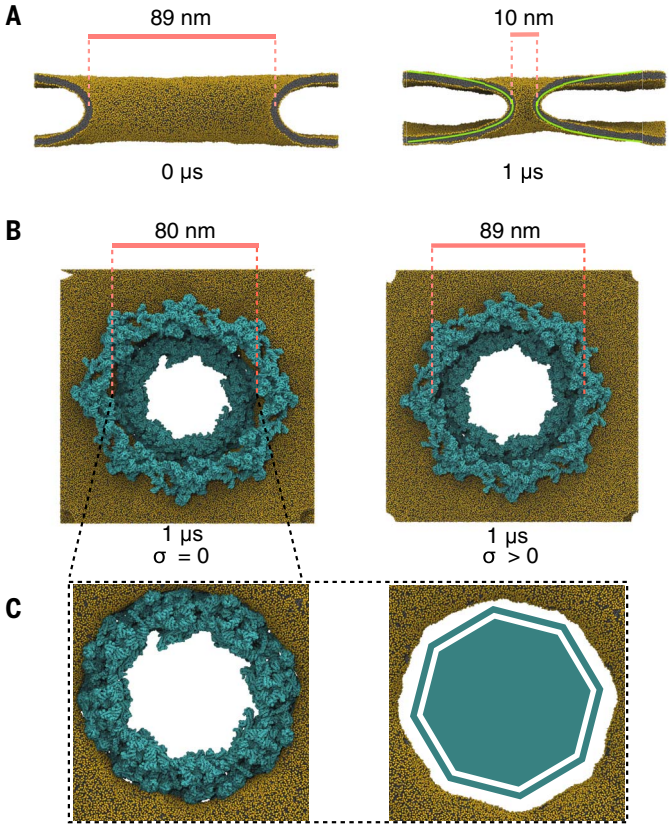


Fig. 4. Dynamics of the NPC from molecular simulations. (**A**) An isolated half-toroidal double-membrane pore shaped initially as in the tomographic structure of the constricted NPC (left) tightens over the course of 1 μs of MD (right) toward the catenoid-like shape (green) predicted by membrane elastic theory. Shown are cuts along the axis of the double-membrane pore with lipid headgroups and tails in gold and gray, respectively. The solvent is not shown. (**B**) The NPC (cyan) widens by ~10% in response to lateral membrane tension (right; ΔP = 2 bar) compared with a zero-tension simulation (left; ΔP = 0). Shown are snapshots of the relaxed structures after 1 μs of MD. (**C**) The membrane fits tightly around the NPC inner ring (cyan, left; ΔP = 0) and forms an octagonally shaped pore (right, NPC not shown).

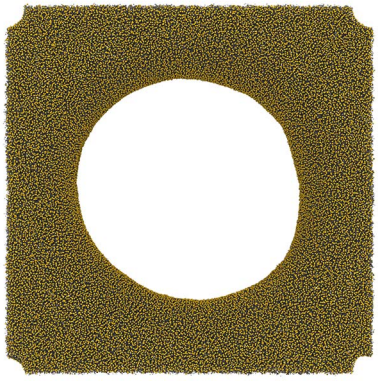
specific physiological conditions (*23*). Our model includes multiple previously unassigned domains and proteins, resolves long-standing ambiguities in alternative NUP assignments, lays out a connectivity map of the protein linkers across the NPC scaffold, maps out the membrane-anchoring motifs, and provides a high-quality basis for further investigations of NPC dynamics and function. Our analysis demonstrates that our model is sufficiently complete for molecular simulations, which in the future could quantitively and predictively describe how the NPC interplays with the nuclear membrane and how it responds to mechanical challenges. The model also provides a more accurate starting point for simulations of nucleocytoplasmic transport by providing the native constraints on the diameter and a more precise mapping of the positions where the FG tails emanate from the scaffold (fig. S22).

How an intricate structure consisting of ~1000 components can be faithfully assembled in the crowded cellular environment is a very intriguing question. Our connectivity map captures the 3D trajectory of linker NUPs through the assembled scaffold. Taken together with previous analysis of NPC assembly (*9*–*13*, *19*), it suggests that the linker NUPs facilitate dedicated spatial organization functions. The connections of NUP93 within individual IR complexes and to the NUP214-complex suggest a role in ensuring isostoichiometric assembly. This finding is consistent with the recent analysis of early NPC biogenesis, suggesting that NUP93 associates isostoichiometrically with the NUP62-complex already during translation in the cytosol (*57*). Thus, the stoichiometric assembly of the NUP62 subcomplex together with NUP205/188-NUP93 heterodimer is likely preassembled away from sites of NPC biogenesis, explaining the importance of the linker for intra-subcomplex interactions. How the spokes form a C2 symmetric interface at the NE plane remains to be addressed.

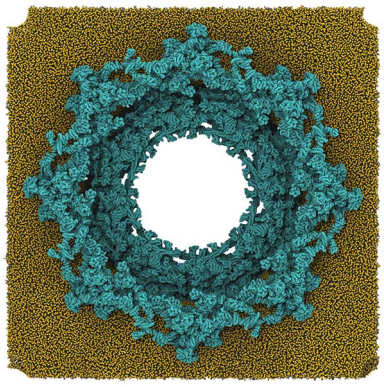
In the IR membrane coat, multiple interactions converge into a distinctive transmembrane interaction hub. We propose that its core is formed by the ALADIN-NDC1 heterodimer at the interface between the outer and inner rings. This transmembrane interaction hub is likely a spatial organizer for two proximate copies of NUP155 within the same spoke that point toward the outer rings and IR, respectively. ALADIN-NDC1 likely further associates with NUP210, which arches between spokes in the NE lumen. The hub also binds NUP35, which connects to NUP155 copies of neighboring spokes, thus facilitating the horizontal, cylindrical oligomerization. Because NUP35 associates with NUP155 early during NPC assembly (*39*), its dimerizing domain appears critical to scaffold its flexible linkers toward neighboring spokes within the IR membrane coat.

The often-emphasized notion that NPCs fuse the INM and ONM or that they stabilize the fusion of the INM and ONM is not necessarily supported by our analysis. Our simulations suggest that the membrane fusion topology per se is stable under certain conditions, relaxing toward a catenoid shape with zero membrane bending energy. Indeed, some species maintain the fusion topology in the absence of NPCs, for example, during semi-closed mitosis in *Drosophila melanogaster* (*58*). Our analysis instead suggests that NPCs stabilize a pore that is wider than in the relaxed, tensionless double-membrane hole. This notion agrees with the ultrastructural analysis of postmitotic NPC assembly, which has revealed that NE holes are formed at small diameters and dilate once NPC subcomplexes are recruited (*59*). These data argue that the membrane shape defines the outline of the NPC scaffold and not vice versa.

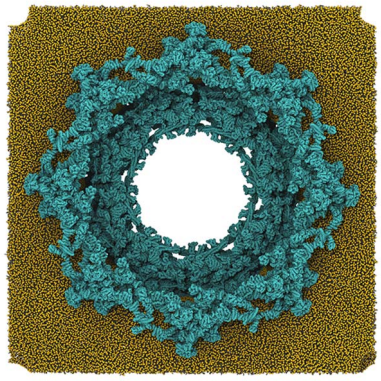
We use AI-based structure prediction programs AlphaFold and RoseTTAfold to model all atomic structures that were used for fitting to the EM maps. Although x-ray and cryo-EM structures were used for validation, no experimental atomic structures were directly incorporated into the model. Predicted atomic structures traditionally exhibited various inaccuracies, limiting their usage for detailed near-atomic model building in low-resolution EM maps. However, AlphaFold and RoseTTAfold have recently demonstrated unprecedented accuracy in predicting structures of monomeric proteins (*26*, *27*, *60*–*65*) and complexes (*34*, *36*, *61*, *66*). They accurately assess their confidence at the level of individual residues and interdomain contacts (*26*, *27*, *61*). Indeed, we could successfully validate our models by comparing them to unpublished crystal structures, cryo-EM maps, and biochemical data. The resulting model of the NPC scaffold is almost complete and exhibits near-atomic-level precision at several interfaces. The model also contains several peripheral NUPs, for example, parts of the NUP214 and NUP358 complexes. Projection of the locally estimated accuracy into an asymmetric unit of the NPC



Movie 1. MD simulation of a half-toroidal double membrane. MD simulation trajectory of an isolated half-toroidal double membrane shaped initially as in the tomographic structure of the constricted NPC. The pore tightens within 1.2 μs of MD (see also Fig. 4B and fig. S23A for the diameter time trace). A top view of lipids is shown. Solvent is omitted for clarity.



Movie 2. MD simulation of an NPC with explicit membrane (α = 1.0). MD simulation of the NPC (cyan) covering ~1.2 μs with α = 1.0 (see also figs. S23 and S27 for the diameter time trace). A top view of the NPC with membrane is shown. Solvent is omitted for clarity.



Movie 3. MD simulation of an NPC with explicit membrane (α = 0.7). MD simulation of the NPC (cyan) covering ~1.2 μs with α = 0.7 (see also Fig. 4B and fig. S23A for the diameter time trace). A top view of the NPC with membrane is shown. Solvent is omitted for clarity.

reveals that the structured regions are generally modeled with good confidence, while linkers and peripheral loops are less well defined (fig. S29). Although the entire EM density for those peripheral NUPs is unlikely to be resolved in the near future owing to their flexibility, the complete model of the human NPC could be within reach by integrating data from complementary techniques that can address flexible proteins, such as super-resolution microscopy, fluorescence resonance energy transfer, and site-specific labeling (*18*).

Thanks to in situ and in cellulo cryo-ET and powerful AI-based prediction (*26*, *27*), intricate structures such as the NPC can now be modeled. Not all subunit or domain combinations that we attempted to model with AI-based structure prediction led to structural models that were consistent with complementary data, emphasizing that experimental structure determination will still be required in the future for cases in which a priori knowledge remains sparse. However, even if AI-based modeling does not yield high-confidence results, the models can still serve as tools for hypothesis generation and subsequent experimental validation.

Materials and methods

Mammalian cell cultivation and subcellular fractionation

Modified human embryonic kidney cells 293 (HEK Flp-In T-Rex) 293 Cell Line, Life Technologies) designed for rapid generation of stably transfected cell lines with a tetracycline-inducible expression system were used as parental cells. The NUP210 CRISPR-knockout line (HEK NUP210Δ) has been previously described (*67*). In general, all cells were maintained in Dulbecco's modified Eagle medium (DMEM) supplemented with 5 g/liter glucose and 10% heat-inactivated fetal bovine serum (FBS, Sigma-Aldrich). HeLa Kyoto cell line was maintained in DMEM medium containing 1 g/liter glucose supplemented with 2 mM L-glutamine. Cells grown close to confluency (~90%) were trypsinized with 0.25% trypsin containing EDTA (Life Technologies) and passaged for further growth. For the preparation of nuclear envelopes, HeLa cells were cultured and subjected to subcellular fraction as described before (*8*, *43*).

Grid preparation

Grids with HeLa nuclear envelopes were prepared exactly as described in (*21*). For the in cellulo work, Au200 R2/1 SiO_2 grids (Quantifoil Micro Tools GmbH) were glow discharged on both sides and sterilized under ultraviolet light. In a six-well cell culture dish, either 250,000 cells per well (for HeLa) or 400,000 cells per well (HEK293) were pipetted onto the grids prewetted with DMEM medium. Cells were left to settle and attach to the grids for 4 hours at 37°C in 5% CO_2. Subsequently, the HeLa or HEK293 grids were plunge-frozen with a Leica EM GP plunger with set chamber environment to 99% humidity and 37°C. Grids were blotted from the backside for 2 s and plunged in liquid ethane-propane mix (37 and 63%) at about −195°C. HEK NUP210Δ grids were washed once with phosphate-buffered saline containing 8% dextran (35.45 kDa) and blotted for 3 s before plunge freezing in −186°C liquid ethane.

Cryo-FIB milling and data acquisition

Plunge-frozen sample grids were FIB-milled on an Aquilos FIB-SEM (Thermo Fisher Scientific) as described before (*22*, *23*). In brief, samples were coated with inorganic platinum (Pt-sputtering). Subsequently, a protective layer of organometallic platinum was deposited for ~20 s using the gas injection system. Cells were then stepwise milled at a 20° angle to a final thickness of ~200 to 250 nm using decreasing ion-beam currents of 1 nA to 50 pA. A final round of Pt-sputtering was applied before unloading the sample.

Cryo–electron tomography and subtomogram averaging of the human NPC from nuclear envelopes

Tilt series were collected with SerialEM as described by Kosinski *et al.* (*19*). The angular coverage of the tilts spanned from −60° to +60°. Ten 8K x 8K frames, per tilt, were collected in the super-resolution mode on a K2 direct electron detector (Gatan Inc.) equipped with an BioQuantum Imaging Filter (GIF). An average total dose of 120 $e^-/Å^2$ per tomogram was used. Five hundred sixteen new tilt series were collected and combined with 101 tilt series reported by Kosinski *et al.* (*19*) leading to a total of 617 tilt series. Incomplete tilt series (missing more than seven tilts in either direction or terminated owing to autofocusing error because of the edge of the grid bar) were discarded. Contrast transfer function (CTF) was determined using CTFFind4 (*68*). Tilt series with large discrepancies in the two defocus values estimated by CTFFind4 were also removed. This resulted in a total of 554 tilt series. Tilt series were manually aligned by tracking gold fiducials in IMOD (*69*). Tilt series were filtered according to accumulated exposure on the basis of parameters described by Kosinski *et al.* (*19*). Tilt series were reconstructed with 3D CTF correction using NovaCTF (*70*).

Subtomograms (7711) containing individual NPCs were extracted, corresponding to 61,688 asymmetric units. The pixel size at the specimen level was 3.37 Å. Tomograms were binned by Fourier cropping 2× (bin2), 4× (bin4), and 8× (bin8), and subtomograms were extracted at each level of binning corresponding to a pixel size of 6.74, 13.48, and 26.96, respectively. Subtomogram averaging was performed on a whole-pore level with the bin8 dataset. Subsequently, asymmetric units were extracted from the aligned pores and averaged as described by Kosinski *et al.* (*19*). Subunits with the center outside of tomogram boundaries were excluded from further processing. The CR, IR, and NR were processed separately, that is, the positions of subunit centers were moved to be in the center of each ring, resulting in three different sets of subtomograms. Rings with the center outside of tomogram boundaries were excluded from further processing. The subtomograms were iteratively aligned first on bin4 and then on bin2 level, and the final alignment was refined on bin1 level. The complete subtomogram averaging and alignment was performed using novaSTA (*71*), the masks necessary for the alignment were created in Dynamo (*72*) and Relion (*73*).

After the bin4 alignment, the quality of each subtomogram was assessed using geometrically restrained classification that is based on the expected geometrical shape of a complete ring. More precisely, the subunits corresponding to one NPC ring should still be part of a ring after the alignment. For each ring and each subunit, the angular distance of its normal vector to all other normal vectors of subunits within the same ring was computed. Subsequently, the distances were averaged (for each NPC separately), and for each subunit, the deviation from the normal vector from the average was computed. The same was done for so-called in-plane vectors, that is, vectors describing the direction from the NPC center to the center of a subunit. The expected/ideal angular distance for the normal vectors is zero, while for the in-plane vectors it is 45°. This analysis was performed only on the rings with at least three subunits that were retained from the initial subtomogram averaging runs. The rings with fewer than three subunits were removed from further processing. The computed deviations were used to identify poorly aligned ring subunits in order to remove them. For CR and SR, all subunits with the normal vector deviation >30° were removed. The deviation of in-plane angles from expected 45° greater than 15° or CR and 10° for SR was used as threshold for additional removal of ring subunits. The threshold values were determined empirically. For CR and SR, the number of subunits left for processing after geometrical cleaning were 31,774 and 35,281, respectively. The geometrical-restrain classification was added to the publicly available novaSTA package (*71*). The final subunit cleaning to remove poor-quality subunits was performed on the final average using the constrained cross-correlation (CCC) value, which was computed between each subtomogram and the reference during the last iteration of alignment. Subtomograms with the worst CCC values were subsequently removed in batches of 1000, as long as the resolution improved. The number of subunits or subtomograms contributing toward the final structure of the CR and SR ring were 21,604 and 30,000, respectively.

In contrast to CR and IR, adding additional tilt series followed by geometric cleaning procedure on NR did not yield any significant improvement in comparison to the dataset reported by Kosinski *et al.* (*19*). Thus, the original map of the NR was used for the presented analysis. The map was created using steps described by Kosinski *et al.* (*19*), and the total number of particles contributing to the final average was 11,112.

Cryo–electron tomography and subtomogram averaging of human NPC in cellulo

Data acquisition was performed on a Titan Krios G2 (for HeLa) or G4 (for HEK) (Thermo Fisher), operating at 300 kV and equipped with Gatan K2 Summit direct electron detector and energy filter as described before (*23*). In brief, tilt series were acquired in dose-fractionation mode at 4k by 4k resolution with a nominal pixel size of 3.37 (HeLa) or 3.45 Å (HEK) using an automated dose-symmetric acquisition scheme (*74*) starting at a given pre-tilt corresponding to the tilt of the FIB-milled lamellae (typically ±13°). Tilt series were acquired with a tilt increment of 3° and a tilt range interval of −50°/+50°, and a total dose per tomogram of 120 to 150 $e^-/Å^2$.

Tilt series preprocessing and tomogram reconstruction was performed as described previously (*22*, *23*). Subtomogram averaging was performed as described before (*22*, *23*). In brief, for the HeLa control dataset, 53 NPCs were extracted from 13 tomograms. For the HEK dataset, 30 control and 43 NUP210Δ NPCs were extracted from 8 control and 14 NUP210Δ tomograms, respectively. Whole pores were aligned using bin8 and bin4 subtomograms with imposed eightfold symmetry. Upon convergence, 280, 150, and 222 subunits were extracted from the control HeLa, control HEK, and HEK NUP210Δ datasets, respectively, and the CR, IR, and NR subunits were further refined independently using bin4 subtomograms. The individual ring subunits were refined without splitting the data into independent half sets to a final resolution of <54 Å (NR) and <48 Å (CR and IR) as estimated by Fourier shell correlation (FSC) using the 0.5 criterion for the HEK datasets. For the HeLa dataset, gold-standard criteria were used to calculate the FSC, which resulted in final resolution of 45 Å (CR and IR) and 53 Å (NR).

Structure determination of human NUP155

The gene encoding human NUP155 (UniProt ID: O75694) was synthesized by GeneArt (Life Technologies) and cloned into a modified pFastBac vector, with a His6 tag and an enhanced green fluorescent protein tag followed by an HRV 3C protease site at the N terminus and a Strep tag at the C terminus. The predicted membrane-binding loop (residues 260 to 273) was deleted to improve the protein stability. The resulting construct was expressed in Sf21 insect cells using the Bac-to-Bac baculovirus expression system (Thermo Fisher Scientific). Sf21 cells were cultured in Sf900III medium (Gibco) at 27°C and infected at a density of 1×10^6 to 2×10^6 cells ml^{-1}. After 48 hours of incubation, cells were collected by centrifugation (3000*g*, 10 min, 27°C), and the pellets were stored at −80°C until purification.

For purification, frozen cell pellet from 100 ml culture was resuspended in 10 ml of buffer containing 20 mM HEPES (pH 7.5), 100 mM NaCl, 1 mM dithiothreitol (DTT), and 0.1 mM phenylmethylsulfonyl fluoride and disrupted by sonication for 5 min using Branson sonifier 250. After removing cell debris by centrifugation (3000*g*, 10 min, 4°C), the supernatant was mixed with 500 µl of Strep-Tactin Sepharose resin (IBA Lifesciences) and incubated at 4°C for 30 min. The resin was washed with 8 ml of buffer containing 20 mM HEPES (pH 7.5), 100 mM NaCl, and 1 mM DTT, and the bound sample was eluted with the same buffer supplemented with 5 mM biotin. The eluted fractions were concentrated with an Amicon Ultra 0.5 ml centrifugal filter (100 kDa molecular weight cut-off, Millipore), mixed with Turbo-3C protease (Sigma-Aldrich), and incubated at 4°C overnight. The sample was then ultracentrifuged (71,680*g*, 15 min, 4°C) and loaded onto Superose 6 Increase 3.2/300 equilibrated with 20 mM HEPES (pH 7.5), 100 mM NaCl, and 1 mM DTT. The peak fractions were aliquoted and stored at −80°C until use.

For the preparation of EM grids, the protein concentration was adjusted to 0.4 mg ml^{-1} in 20 mM HEPES (pH 7.5), 100 mM NaCl, 1 mM DTT, and 0.001% dodecyl maltoside. The diluted sample was then applied onto a freshly glow-discharged UltrAuFoil R0.6/1.0 gold grids (300 mesh, Quantifoil), blotted for 6 s at 4°C in 100% humidity, and plunge-frozen in liquid ethane using Vitrobot Mark IV. Cryo-EM data were collected on a Titan Krios G4 microscope (Thermo Scientific) operated at 300 kV, equipped with a E-CFEG, a Falon 4 direct electron detector (Thermo Scientific), and a Selectris X energy filter (Thermo Scientific) operated with a slit width of 10 eV. Automated data acquisition was performed with the EPU software at a nominal magnification of ×165,000, corresponding to a

pixel size of 0.730 Å per pixel. Movies were acquired at a dose rate of 4.95 electrons per pixel per second, and a total dose of 50 $e^-/Å^2$, resulting in EER movies consisting of 1407 frames. In total, 6430 movies were acquired with a defocus range of −1.0 to −2.0 µm. Acquired images were first processed in cryoSPARC (*75*), and selected particles were further processed in RELION-3.1 using csparc2star.py script in UCSF pyem (*76*) for transfer of particles. To analyze the conformational flexibility, multibody refinement (*77*) was performed on the consensus map. The details of data processing are summarized in fig. S8.

The model of human NUP155 was manually built in *Coot* (*78*), using the crystal structure of Nup170, a homolog of NUP155, from *C. thermophilum* (PDB ID: 5HAX) as a starting model. Secondary structure prediction from PSIPRED (*79*) and multiple sequence alignment were used to facilitate the model building. The model was iteratively refined using phenix.real_space_refine (*80*). Figures were prepared with UCSF Chimera (*81*), UCSF ChimeraX (*82*), and CueMol (http://www.cuemol.org/).

Proximity labeling using BioID

BioID analysis of ALADIN was done as previously described (*83*). In brief, ALADIN was BirA-tagged and overexpressed in Hek293 Flp-In Trex cells. Quantitative mass spectrometry was done in four biological replicates and in comparison to control cells expressing BirA-tagged NLS-NES-Dendra that resides within the central channel.

Structural modeling of NUPs and NPC subcomplexes

The structures of all individual NUPs and selected subcomplexes were modeled using AlphaFold (*26*) or downloaded from AlphaFold Database (*60*). The models of monomeric proteins (NUP155, NUP133, NUP107, NUP93, NUP205, NUP188, NUP160, NUP358, and ELYS) were download from AlphaFold Database (*60*). To model subcomplexes or their parts around the interfaces (NUP62-NUP54-NUP58, NUP205-NUP93, NUP188-NUP93, NUP155-NUP35, NUP93-NUP35, NDC1-ALADIN, NUP35 homodimer, NUP85-SEH1-NUP43, NUP160-NUP96-SEC13, NUP160-NUP37, NUP133-NUP107, NUP96-NUP107, NUP160-NUP96-NUP85, NUP214-NUP62-NUP88, and NUP88-NUP98), we used the AlphaFold version modified for modeling complexes, available through ColabFold (*35*), with all parameters set to default except for the max_recycles parameter, which was set to between 12 and 48, depending on the subcomplex. For NUP210, we first built the initial full-length using RoseTTAfold (*27*), as AlphaFold did not provide a full-length model fitting well into the EM density map. After fitting the model into the EM maps as a rigid body (see below), we used AlphaFold to model successive monomeric and homodimeric fragments of NUP210, superposed them onto the fitted RoseTTAfold model, and refined the fits. The quality of the AlphaFold models was first assessed by the scores provided by the authors—the predicted local distance difference test (pLDDT), which predicts the local accuracy, and predicted aligned error, which assesses the packing between domains and protein chains. In addition, we validated the models by comparing to structures not used for modeling, structures published in the accompanying paper (*32*), fits to the cryo-ET maps, and previously published biochemical data (figs. S5 to S7, S10, S11, and S14).

Systematic fitting of atomic structures to cryo-ET maps

We used the previously published procedure for systematic fitting (*8*, *19*, *21*, *22*, *25*, *37*, *84*) to both locate the atomic structures in the cryo-ET maps and validate the AlphaFold models. Before fitting, all the high-resolution structures were filtered to between 10 and 15 Å. The resulting simulated model maps were subsequently fitted into individual ring segments of cryo-ET maps by global fitting as implemented in UCSF Chimera (*82*) using scripts in Assembline (*37*). The maps used for fitting excluded nuclear envelope density in order to eliminate the possibility of fits overlapping with the membrane. All fitting runs were performed using 100,000 random initial placements, with the requirement of at least 30 to 60% (depending on the size of the structure) of the simulated model map to be covered by the cryo-ET density envelope defined at a low threshold. For each fitted model, this procedure resulted in ~1000 to 20,000 fits with nonredundant conformations upon clustering. The cross-correlation about the mean (cam score, equivalent to Pearson correlation) score from UCSF Chimera (*81*) was used as a fitting metric for each atomic structure, similarly to our previously published works. The statistical significance of every fitted model was evaluated as a *P* value derived from the cam scores. The calculation of *P* values was performed by first transforming the cross-correlation scores to *z*-scores (Fisher's *z*-transform) and centering, from which subsequently two-sided *P* values were computed using standard deviation derived from an empirical null distribution [based on all obtained nonredundant fits and fitted using fdrtool (*85*) R-package]. Finally, the *P* values were corrected for multiple testing with Benjamini-Hochberg procedure (*86*).

Modeling of the human NPC scaffold

To assemble the models of the entire NPC scaffold based on the constricted and dilated cryo-ET maps, we used our integrative modeling software Assembline (*37*), which is based on Integrative Modeling Platform (IMP) (*87*) version 2.15 and Python Modeling Interface (PMI) (*88*). First, we built the model of the constricted NPC owing to its higher resolution. The AlphaFold models of NUP domains and subcomplexes already present in our previous human NPC models (*19*, *21*) were placed in the map by superposing them onto the published models. The remaining domains and subcomplexes added in this work (NUP358, NUP35, NUP93 in the outer rings, NDC, ALADIN, and the NUP214 complex) were placed using systematic fitting (as above) and global optimization procedure of Assembline. In addition to using models of subcomplexes as rigid bodies for fitting, several inter-subunit interfaces were restrained by elastic distance network derived from ColabFold models overlapping with and bridging already fitted models. During the refinement, the structures were used as rigid bodies and simultaneously represented at two resolutions: in Cα-only representation and a coarse-grained representation, in which each 10-residue stretch was converted to a bead. The 10-residue bead representation was used for all restraints to increase computational efficiency except for the domain connectivity restraints, for which the Cα-only representation was used. The flexible protein linkers between the domains were added as chains of one-residue beads. The entire structure was optimized using the refinement step of Assembline to optimize the fit to the map, minimize steric clashes, and ensure connectivity of the protein linkers. The scoring function for the refinement comprised the EM fit restraint; clash score (SoftSpherePairScore of IMP); connectivity distance between domains neighboring in sequence; a term preventing overlap of the protein mass with the nuclear envelope; a restraint promoting the membrane-binding loops of NUP133, NUP160, and NUP155 to interact with the envelope implemented using MapDistanceTransform of IMP [predicted by similarity to known or predicted ALPS motifs *X. laevis* and *S. cerevisiae* homologs (*6*, *21*, *89*)]; and elastic network restraints derived from the subcomplexes modeled with AlphaFold/ColabFold. The final atomic structures were generated from the refined models by back-mapping the coarse-grained representation to the original AlphaFold atomic models. The conformation of the linkers was further optimized using Modeller (*90*) and Isolde (*91*). The stereochemistry of the final model was optimized using steepest descent minimization in GROMACS (*92*).

The model of the dilated NPC was built by fitting the asymmetric units of the individual cytoplasmic, inner, and nuclear rings of the constricted NPC model to the dilated cryo-ET maps and refining the fits with Assembline. The refinement procedure was performed as above.

To calculate the percentage of the molecular weight of the full NPC and the NPC scaffold covered by the new and the old models, we defined the full NPC as being composed of the following 32 NUPs, with the stoichiometry indicated in the parentheses: NUP160 (32), NUP96 (32), NUP85 (32), SEH1 (32), SEC13 (32), NUP107 (32), NUP133 (32), NUP358 (40), NUP43 (32), ELYS (16), NUP37 (32), NUP188 (16), NUP205 (40), NUP155 (48), NUP93 (56), NUP35 (32), NUP62 (48), NUP54 (32), NUP58 (32), NUP88 (16), NUP214 (16), NUP98 (48), NDC1 (16), NUP210 (64), and ALADIN (16), POM121 (32), TPR (32), NUP153 (32), NUP50 (16), CG1 (8), DDX19 (16), and GLE1 (8). The scaffold NPC was defined as being composed of 25 NUPs: NUP160, NUP96, NUP85, SEH1, SEC13, NUP107, NUP133, NUP358, NUP43, ELYS, NUP37, NUP188, NUP205, NUP155, NUP93, NUP35, NUP62, NUP54, NUP58, NUP88, NUP214, NUP98, NDC1, NUP210, and ALADIN. The stoichiometry for the scaffold was the same as for the full NPC with exception of NUP214 complex for which only one copy was counted, as the second copy is not clearly visible in the EM density. Note that for some nucleoporins, like NUP98 or POM121, the exact stoichiometry is still uncertain. The coiled-coil domains of the peripheral NUPs of the NUP214 complex and the α-solenoid domain of NUP358 were included in the scaffold. The FG regions were excluded. These definitions resulted in the molecular weight of 119 MDa for the full NPC and 76 MDa for the scaffold. The scaffold diameters were described by two distances between the opposite spokes: the membrane-to-membrane distance and the distance between ferredoxin-like domains of NUP54 at the residue 220. Figures were produced using UCSF ChimeraX (*82*).

Molecular dynamics (MD) simulations

We performed MD simulations of half-toroidal membrane pores in isolation and including the hNPC scaffold. In the following section, we describe the setup of the simulation models, the relevant MD parameters, and the analysis of the MD trajectories.

Membrane model

First, a 30 nm by 30 nm coarse-grained POPC lipid bilayer patch was generated using *insane.py* (*93*, *94*). The bilayer was placed in a periodic simulation box, solvated on both sides, energy minimized, and simulated for 100 ns using standard MD parameters, as noted below.

Then, half-toroidal membrane pores were constructed with the BUMpy software (*93*) using this initial flat bilayer as membrane input. The following command line flags were used when running bumpy: -s double_bilayer_cylinder -z 10 -g l_cylinder:10 r_cylinder:430 r_junction:120 l_flat:1400 (see bumpy documentation). The resulting membranes coincided reasonably with the cryo-ET density of the double-membrane pore and allowed us to place the membrane-anchoring motifs of the NPC model into the membrane.

Two carbon nanotube porins (CNTPs) were inserted into the membrane in the corners of the simulation box distant from the NPC (see, e.g., fig. S19A). The CNTPs with a length of 3.6 nm and a diameter of 14.7 nm enable water transfer in and out of the otherwise disconnected luminal volume, as is required for membrane-mechanical equilibration. Without CNTPs, the luminal volume would be effectively fixed and, as a result, changes in the membrane shape during MD simulations without NPC scaffold would induce artifactual membrane buckling. The CNTPs were built according to previous work (*93*–*95*). The outermost carbon rings at either CNTP end consisted of polar SNda beads for stable membrane embedding (*93*). To stiffen the wide CNTPs, the improper dihedral force constant was increased to 1000 kJ mol^{-1} rad^{-2}. The CNTP parameters were otherwise set as previously reported (*95*, *96*). The code to generate CNTP models and the parameters for simulations are available at: https://github.com/bio-phys/cnt-martini (*95*). The CNTPs were embedded in the flat patch of the NPC membranes away from the NPC. Lipids within 8 Å of the CNTPs or inside their circumference were removed.

NPC scaffold model

The MD simulation model of the NPC included the entire scaffold (see table S4 and fig. S19 for a summary of the hNPC simulation model) except for the disordered FG-NUP C- and N-terminal tails. For simplicity and to limit the system size, we also excluded the NUP210 glycoprotein in the nuclear envelope lumen. Otherwise, the models were complete as described above.

Each protein chain was coarse-grained individually using *martinize.py* as follows. All chain termini were uncharged and otherwise default protonation states were used. Secondary structure restraints were assigned according to DSSP (*97*). The tertiary structure of each protein chain was maintained by an elastic network using the recommended default settings with a cutoff R_c of 0.9 nm and a force constant k of 500 kJ mol^{-1} nm^{-2}. For each protein chain, the ElNeDyn2.2 protein force field was used in conjunction with the Martini 2.2 force field (*53*, *54*). Simulations were performed with the default protein-protein interaction (α = 1.0; results shown in the supplementary materials) and with protein-protein interactions scaled relative to protein-solvent interactions with α = 0.7 (*98*) (results shown in the main text) to correct for the effect of reportedly overestimated nonbonded interactions (*99*). This procedure used the *martinize.py* script and was wrapped in custom python code to automatically generate the structures for each protein chain with the aforementioned parameters. To enable easier handling of the large number of protein chains, each protein chain was assigned a unique segid. Importantly, with this MD simulation model, all protein-protein interactions between distinct chains could dissociate and new interactions could form in principle, and the structure of linker regions could relax.

All individually coarse-grained protein chains were then merged into one PDB structure file. The resulting coarse-grained NPC scaffold model was centered within the half-toroidal membrane pore model containing the CNTPs described above. Any lipids within 8 Å of any bead of the scaffold proteins in the initial assembly were removed.

Solvation

All systems were solvated with coarse-grained water containing 10% anti-freeze WF particles and Na^+ ions to neutralize the system using standard GROMACS tools. All systems simulated in this study are listed in table S5.

MD simulations

All molecular dynamics simulations were performed using the GROMACS software package and the coarse-grained Martini force field v2.2 (*53*, *92*, *100*). Each system was first steepest-descent energy minimized using a soft-core potential to remove steric clashes in the initial model. The systems were then equilibrated in an NPT ensemble with semisotropic pressure coupling first for 2.5 ns with a 5-fs timestep and then for 100 ns with a 15-fs timestep with position restraints on the protein backbone beads with a force constant of 1000 kJ mol^{-1} nm^{-2}, maintaining a temperature of 310 K and pressure of 1 bar using the Berendsen barostat and velocity rescaling thermostat (*101*, *102*). Characteristic coupling times of 12 and 1 ps were used, respectively. During production simulations, the Parrinello-Rahman barostat was used (*103*).

The Verlet neighbor search algorithm was used to update the neighbor list, with the length and update frequency being automatically determined. Lennard-Jones and Coulomb forces were cut off at 1.1 nm, with the potential shifted to zero using the Verlet-shift potential modifier. A 15-fs timestep was used in all production simulations. Production simulations were performed for ~1.2 μs each.

Membrane tension

To apply lateral tension on the double-membrane structure, an anisotropic pressure tensor was used with an out-of-plane pressure of $P_\perp = 2\Delta P/3$ and an in-plane pressure of $P_\parallel = p - \Delta P/3$, with $p = 1$ bar. This results in a traceless lateral strain

$S = diag(-\Delta P/3, -\Delta P/3, 2\Delta P/3)$ where $\Delta P \equiv P_\perp - P_\parallel$. The resulting tension on the double-membrane system is $\sigma = (P_\perp - P_\parallel)L_z = \Delta P L_z$ with L_z the box height. To allow for gradual equilibration under tension, ΔP was increased in steps of 1 bar until reaching the target value (see table S1).

Analysis of MD simulations

Images and movies were generated using VMD (*104*) and time series were analyzed using the MDAnalysis library (*105*). To monitor conformational changes, we calculated the root-mean-square distance (RMSD) from the starting structure using the qcprot RMSD alignment algorithm implemented in MDAnalysis (*105*). The RMSD was calculated every 1.5 ns for the backbone (BB) beads with respect to the rigid-body aligned initial structure. In addition to the individual protein chains, we analyzed in this way the β-propeller present in the three nucleoporins NUP133 (residues 1 to 480), NUP155 (residues 1 to 500), and NUP160 (residues 1 to 500); as well as the respective alpha solenoid domain NUP133 (residues 500 to end), and NUP155, and NUP160 (residues 507 to end); and each of the eight spokes as a whole. In the RMSD analysis, averages and standard deviations were calculated across the eight spokes or across equivalent protein copies in the NPC scaffold, respectively.

During the MD simulations, the diameter of the NPC membrane pore was determined by least-square fitting the center and radius of a circle in the xy-plane to the membrane center (C4A and C4B lipid beads). The fit was performed at the narrowest region of the half-toroidal membrane pore.

Possible limitations

We note that the time scale currently accessible to MD simulations is too short to fully recapitulate the complete NPC dilation and constriction processes, including the large-scale NPC structural rearrangements. We also note that the elastic network on proteins of the Martini model restricts internal conformational changes, which might be required for larger-scale NPC dilation. The coarse-grained interaction model may also weaken some protein-protein interactions and strengthen others. Finally, we expect that the missing FG mesh in the MD model contributes to the compaction of the NPC scaffold seen in the MD simulations, acting on top of the mechanical tension in the widened double-membrane pore (supplementary text).

REFERENCES AND NOTES

1. A. Ibarra, M. W. Hetzer, Nuclear pore proteins and the control of genome functions. *Genes Dev.* **29**, 337–349 (2015). doi: 10.1101/gad.256495.114; pmid: 25691464
2. M. Raices, M. A. D'Angelo, Nuclear pore complexes and regulation of gene expression. *Curr. Opin. Cell Biol.* **46**, 26–32 (2017). doi: 10.1016/j.ceb.2016.12.006; pmid: 28088069
3. C. Strambio-De-Castillia, M. Niepel, M. P. Rout, The nuclear pore complex: Bridging nuclear transport and gene regulation. *Nat. Rev. Mol. Cell Biol.* **11**, 490–501 (2010). doi: 10.1038/nrm2928; pmid: 20571586
4. P. De Magistris, W. Antonin, The dynamic nature of the nuclear envelope. *Curr. Biol.* **28**, R487–R497 (2018). doi: 10.1016/j.cub.2018.01.073; pmid: 29689232
5. B. Hampoelz, A. Andres-Pons, P. Kastritis, M. Beck, Structure and assembly of the nuclear pore complex. *Annu. Rev. Biophys.* **48**, 515–536 (2019). doi: 10.1146/annurev-biophys-052118-115308; pmid: 30943044
6. D. H. Lin, A. Hoelz, The structure of the nuclear pore complex (an update). *Annu. Rev. Biochem.* **88**, 725–783 (2019). doi: 10.1146/annurev-biochem-062917-011901; pmid: 30883195
7. S. Siniossoglou *et al.*, Structure and assembly of the Nup84p complex. *J. Cell Biol.* **149**, 41–54 (2000). doi: 10.1083/jcb.149.1.41; pmid: 10747086
8. K. H. Bui *et al.*, Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* **155**, 1233–1243 (2013). doi: 10.1016/j.cell.2013.10.055; pmid: 24315095
9. M. Beck, E. Hurt, The nuclear pore complex: Understanding its function through structural insight. *Nat. Rev. Mol. Cell Biol.* **18**, 73–89 (2017). doi: 10.1038/nrm.2016.147; pmid: 27999437
10. J. Fischer, R. Teimer, S. Amlacher, R. Kunze, E. Hurt, Linker Nups connect the nuclear pore complex inner ring with the outer ring and transport channel. *Nat. Struct. Mol. Biol.* **22**, 774–781 (2015). doi: 10.1038/nsmb.3084; pmid: 26344569
11. D. H. Lin *et al.*, Architecture of the symmetric core of the nuclear pore. *Science* **352**, aaf1015 (2016). doi: 10.1126/science.aaf1015; pmid: 27081075
12. S. J. Kim *et al.*, Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475–482 (2018). doi: 10.1038/nature26003; pmid: 29539637
13. T. Stuwe *et al.*, Architecture of the fungal nuclear pore inner ring complex. *Science* **350**, 56–64 (2015). doi: 10.1126/science.aac9176; pmid: 26316600
14. S. G. Brohawn, T. U. Schwartz, Molecular architecture of the Nup84-Nup145C-Sec13 edge element in the nuclear pore complex lattice. *Nat. Struct. Mol. Biol.* **16**, 1173–1177 (2009). doi: 10.1038/nsmb.1713; pmid: 19855394
15. S. O. Obado *et al.*, Interactome mapping reveals the evolutionary history of the nuclear pore complex. *PLOS Biol.* **14**, e1002365 (2016). doi: 10.1371/journal.pbio.1002365; pmid: 26891179
16. A. G. Floch, B. Palancade, V. Doye, Fifty years of nuclear pores and nucleocytoplasmic transport studies: Multiple tools revealing complex rules. *Methods Cell Biol.* **122**, 1–40 (2014). doi: 10.1016/B978-0-12-417160-2.00001-1; pmid: 24857723
17. D. Görlich, U. Kutay, Transport between the cell nucleus and the cytoplasm. *Annu. Rev. Cell Dev. Biol.* **15**, 607–660 (1999). doi: 10.1146/annurev.cellbio.15.1.607; pmid: 10611974
18. E. A. Lemke, The multiple faces of disordered nucleoporins. *J. Mol. Biol.* **428**, 2011–2024 (2016). doi: 10.1016/j.jmb.2016.01.002; pmid: 26791761
19. J. Kosinski *et al.*, Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352**, 363–365 (2016). doi: 10.1126/science.aaf0643; pmid: 27081072
20. A. P. Schuller *et al.*, The cellular environment shapes the nuclear pore complex architecture. *Nature* **598**, 667–671 (2021). doi: 10.1038/s41586-021-03985-3; pmid: 34646014
21. A. von Appen *et al.*, *In situ* structural analysis of the human nuclear pore complex. *Nature* **526**, 140–143 (2015). doi: 10.1038/nature15381; pmid: 26416747
22. M. Allegretti *et al.*, In-cell architecture of the nuclear pore and snapshots of its turnover. *Nature* **586**, 796–800 (2020). doi: 10.1038/s41586-020-2670-5; pmid: 32879490
23. C. E. Zimmerli *et al.*, Nuclear pores dilate and constrict in cellulo. *Science* **374**, eabd9776 (2021). doi: 10.1126/science.abd9776; pmid: 34762489
24. M. Beck, S. Mosalaganti, J. Kosinski, From the resolution revolution to evolution: Structural insights into the evolutionary relationships between vesicle coats and the nuclear pore. *Curr. Opin. Struct. Biol.* **52**, 32–40 (2018). doi: 10.1016/j.sbi.2018.07.012; pmid: 30103204
25. S. Mosalaganti *et al.*, In situ architecture of the algal nuclear pore complex. *Nat. Commun.* **9**, 2361 (2018). doi: 10.1038/s41467-018-04739-y; pmid: 29915221
26. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). doi: 10.1038/s41586-021-03819-2; pmid: 34265844
27. M. Baek *et al.*, Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021). doi: 10.1126/science.abj8754; pmid: 34282049
28. T. Maimon, N. Elad, I. Dahan, O. Medalia, The human nuclear pore complex as revealed by cryo-electron tomography. *Structure* **20**, 998–1006 (2012). doi: 10.1016/j.str.2012.03.025; pmid: 22632834
29. J. Mahamid *et al.*, Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* **351**, 969–972 (2016). doi: 10.1126/science.aad8857; pmid: 26917770
30. V. Zila *et al.*, Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell* **184**, 1032–1046.e18 (2021). doi: 10.1016/j.cell.2021.01.025; pmid: 33571428
31. C. J. Bley *et al.*, Architecture of the cytoplasmic face of the nuclear pore. *Science* **376**, eabm9129 (2022). doi: 10.1126/science.abm9129
32. S. Petrovic *et al.*, Architecture of the linker-scaffold in the nuclear pore. *Science* **376**, eabm9798 (2022). doi: 10.1126/science.abm9798
33. G. Huang *et al.*, Structure of the cytoplasmic ring of the *Xenopus laevis* nuclear pore complex by cryo-electron microscopy single particle analysis. *Cell Res.* **30**, 520–531 (2020). doi: 10.1038/s41422-020-0319-4; pmid: 32376910
34. R. Evans *et al.*, Protein complex prediction with AlphaFold-Multimer. bioRxiv 2021.10.04.463034 [Preprint] (2021). https://doi.org/10.1101/2021.10.04.463034.
35. M. Mirdita *et al.*, ColabFold - Making protein folding accessible to all. bioRxiv 2021.08.15.456425 [Preprint] (2021). https://doi.org/10.1101/2021.08.15.456425.
36. I. R. Humphreys *et al.*, Computed structures of core eukaryotic protein complexes. *Science* **374**, eabm4805 (2021). doi: 10.1126/science.abm4805; pmid: 34762488
37. V. Rantos, K. Karius, J. Kosinski, Integrative structural modeling of macromolecular complexes using Assembline. *Nat. Protoc.* **17**, 152–176 (2022). doi: 10.1038/s41596-021-00640-z; pmid: 34845384
38. N. Schrader *et al.*, Structural basis of the nic96 subcomplex organization in the nuclear pore channel. *Mol. Cell* **29**, 46–55 (2008). doi: 10.1016/j.molcel.2007.10.022; pmid: 18206968
39. B. Vollmer *et al.*, Dimerization and direct membrane interaction of Nup53 contribute to nuclear pore complex assembly. *EMBO J.* **31**, 4072–4084 (2012). doi: 10.1038/emboj.2012.256; pmid: 22960634
40. J. Mansfeld *et al.*, The conserved transmembrane nucleoporin NDC1 is required for nuclear pore complex assembly in vertebrate cells. *Mol. Cell* **22**, 93–103 (2006). doi: 10.1016/j.molcel.2006.02.015; pmid: 16600873
41. Y. Yamazumi *et al.*, The transmembrane nucleoporin NDC1 is required for targeting of ALADIN to nuclear pore complexes. *Biochem. Biophys. Res. Commun.* **389**, 100–104 (2009). doi: 10.1016/j.bbrc.2009.08.096; pmid: 19703420
42. B. Kind, K. Koehler, M. Lorenz, A. Huebner, The nuclear pore complex protein ALADIN is anchored via NDC1 but not via POM121 and GP210 in the nuclear envelope. *Biochem. Biophys. Res. Commun.* **390**, 205–210 (2009). doi: 10.1016/j.bbrc.2009.09.080; pmid: 19782045
43. A. Ori *et al.*, Cell type-specific nuclear pores: A case in point for context-dependent stoichiometry of molecular machines. *Mol. Syst. Biol.* **9**, 648 (2013). doi: 10.1038/msb.2013.4; pmid: 23511206
44. N. Eisenhardt, J. Redolfi, W. Antonin, Interaction of Nup53 with Ndc1 and Nup155 is required for nuclear pore complex assembly. *J. Cell Sci.* **127**, 908–921 (2014). pmid: 24363447
45. D. H. Lin *et al.*, Structural and functional analysis of mRNA export regulation by the nuclear pore complex. *Nat. Commun.* **9**, 2319 (2018). doi: 10.1038/s41467-018-04459-3; pmid: 29899397
46. J. M. Cronshaw, M. J. Matunis, The nuclear pore complex protein ALADIN is mislocalized in triple A syndrome. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5823–5827 (2003). doi: 10.1073/pnas.1031047100; pmid: 12730363
47. A. Huebner *et al.*, The triple A syndrome is due to mutations in ALADIN, a novel member of the nuclear pore complex. *Endocr. Res.* **30**, 891–899 (2004). doi: 10.1081/ERC-200044138; pmid: 15666842

48. A.-R. Cho *et al*., Tissue-specific expression and subcellular localization of ALADIN, the absence of which causes human triple A syndrome. *Exp. Mol. Med.* **41**, 381–386 (2009). doi: 10.3858/emm.2009.41.6.043; pmid: 19322026
49. P. Upla *et al*., Molecular architecture of the major membrane ring component of the nuclear pore complex. *Structure* **25**, 434–445 (2017). doi: 10.1016/j.str.2017.01.006; pmid: 28162953
50. Y. Zhang *et al*., Molecular architecture of the luminal ring of the *Xenopus laevis* nuclear pore complex. *Cell Res.* **30**, 532–540 (2020). doi: 10.1038/s41422-020-0320-y; pmid: 32367042
51. E. Onischenko, L. H. Stanton, A. S. Madrid, T. Kieselbach, K. Weis, Role of the Ndc1 interaction network in yeast nuclear pore complex assembly and maintenance. *J. Cell Biol.* **185**, 475–491 (2009). doi: 10.1083/jcb.200810030; pmid: 19414609
52. D. Devos *et al*., Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLOS Biol.* **2**, e380 (2004). doi: 10.1371/journal.pbio.0020380; pmid: 15523559
53. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. de Vries, The MARTINI force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007). doi: 10.1021/jp071097f; pmid: 17569554
54. X. Periole, M. Cavalli, S.-J. Marrink, M. A. Ceruso, Combining an elastic network with a coarse-grained molecular force field: Structure, dynamics, and intermolecular recognition. *J. Chem. Theory Comput.* **5**, 2531–2543 (2009). doi: 10.1021/ct9002114; pmid: 26616630
55. J. G. Gall, Octagonal nuclear pores. *J. Cell Biol.* **32**, 391–399 (1967). doi: 10.1083/jcb.32.2.391; pmid: 10976230
56. C. Wolf, M. R. K. Mofrad, On the octagonal structure of the nuclear pore complex: Insights from coarse-grained models. *Biophys. J.* **95**, 2073–2085 (2008). doi: 10.1529/biophysj.108.130336; pmid: 18487299
57. M. Seidel *et al*., Co-translational assembly orchestrates competing biogenesis pathways. *Nat. Commun.* **13**, 1224 (2022). doi: 10.1038/s41467-022-28878-5; pmid: 35264577
58. J. P. Stafstrom, L. A. Staehelin, Dynamics of the nuclear envelope and of nuclear pore complexes during mitosis in the *Drosophila* embryo. *Eur. J. Cell Biol.* **34**, 179–189 (1984). pmid: 6428889
59. S. Otsuka *et al*., Postmitotic nuclear pore assembly proceeds by radial dilation of small membrane openings. *Nat. Struct. Mol. Biol.* **25**, 21–28 (2018). doi: 10.1038/s41594-017-0001-9; pmid: 29323269
60. K. Tunyasuvunakool *et al*., Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021). doi: 10.1038/s41586-021-03828-1; pmid: 34293799
61. M. Akdel *et al*., A structural biology community assessment of AlphaFold 2 applications. bioRxiv 2021.09.26.461876 [Preprint] (2021). https://doi.org/10.1101/2021.09.26.461876.
62. G. Masrati *et al*., Integrative structural biology in the era of accurate structure prediction. *J. Mol. Biol.* **433**, 167127 (2021). doi: 10.1016/j.jmb.2021.167127; pmid: 34224746
63. J. Pereira *et al*., High-accuracy protein structure prediction in CASP14. *Proteins* **89**, 1687–1699 (2021). doi: 10.1002/prot.26171; pmid: 34218458
64. V. Ruiz-Serra *et al*., Assessing the accuracy of contact and distance predictions in CASP14. *Proteins* **89**, 1888–1900 (2021). doi: 10.1002/prot.26248; pmid: 34595772
65. J. Jumper *et al*., Applying and improving AlphaFold at CASP14. *Proteins* **89**, 1711–1721 (2021). doi: 10.1002/prot.26257; pmid: 34599769
66. P. Bryant, G. Pozzati, A. Elofsson, Improved prediction of protein-protein interactions using AlphaFold2. *Nat. Commun.* **13**, 1265 (2022). doi: 10.1038/s41467-022-28865-w; pmid: 35273146
67. K. Buczak, "Spatial proteomics: from tissue organization to protein function," thesis, Universität Heidelberg (2019). doi: 10.11588/HEIDOK.00025909
68. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015). doi: 10.1016/j.jsb.2015.08.008; pmid: 26278980
69. D. N. Mastronarde, S. R. Held, Automated tilt series alignment and tomographic reconstruction in IMOD. *J. Struct. Biol.* **197**, 102–113 (2017). doi: 10.1016/j.jsb.2016.07.011; pmid: 27444392
70. B. Turoňová, F. K. M. Schur, W. Wan, J. A. G. Briggs, Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4Å. *J. Struct. Biol.* **199**, 187–195 (2017). doi: 10.1016/j.jsb.2017.07.007; pmid: 28743638
71. B. Turonova, turonova/novaSTA: Advanced particle analysis, version 1.1, Zenodo (2022); https://doi.org/10.5281/zenodo.5921012.doi: 10.5281/zenodo.5921012
72. D. Castaño-Díez, M. Kudryashev, M. Arheit, H. Stahlberg, Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *J. Struct. Biol.* **178**, 139–151 (2012). doi: 10.1016/j.jsb.2011.12.017; pmid: 22245546
73. 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
74. W. J. H. Hagen, W. Wan, J. A. G. Briggs, Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191–198 (2017). doi: 10.1016/j.jsb.2016.06.007; pmid: 27313000
75. 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
76. D. Asarnow, E. Palovcak, Y. Cheng, asarnow/pyem: UCSF pyem v0.5, Zenodo (2019); https://doi.org/10.5281/zenodo.
77. T. Nakane, D. Kimanius, E. Lindahl, S. H. H. W. Scheres, Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7**, e36861 (2018). doi: 10.7554/eLife.36861; pmid: 29856314
78. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010). doi: 10.1107/S0907444910007493; pmid: 20383002
79. L. J. McGuffin, K. Bryson, D. T. Jones, The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000). doi: 10.1093/bioinformatics/16.4.404; pmid: 10869041
80. P. V. Afonine *et al*., Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.* **74**, 531–544 (2018). doi: 10.1107/S2059798318006551; pmid: 29872004
81. E. F. Pettersen *et al*., UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254
82. T. D. Goddard *et al*., UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018). doi: 10.1002/pro.3235; pmid: 28710774
83. M.-T. Mackmull *et al*., Landscape of nuclear transport receptor cargo specificity. *Mol. Syst. Biol.* **13**, 962 (2017). doi: 10.15252/msb.20177608; pmid: 29254951
84. M. I. Dauden *et al*., Architecture of the yeast Elongator complex. *EMBO Rep.* **18**, 264–279 (2017). doi: 10.15252/embr.201643353; pmid: 27974378
85. K. Strimmer, fdrtool: A versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* **24**, 1461–1462 (2008). doi: 10.1093/bioinformatics/btn209; pmid: 18441000
86. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995). doi: 10.1111/j.2517-6161.1995.tb02031.x
87. B. Webb *et al*., Integrative structure modeling with the Integrative Modeling Platform. *Protein Sci.* **27**, 245–258 (2018). doi: 10.1002/pro.3311; pmid: 28960548
88. D. Saltzberg *et al*., Modeling biological complexes using Integrative Modeling Platform. *Methods Mol. Biol.* **2022**, 353–377 (2019). doi: 10.1007/978-1-4939-9608-7_15; pmid: 31396911
89. B. Vollmer *et al*., Nup153 recruits the Nup107-160 complex to the inner nuclear membrane for interphasic nuclear pore complex assembly. *Dev. Cell* **33**, 717–728 (2015). doi: 10.1016/j.devcel.2015.04.027; pmid: 26051542
90. B. Webb, A. Sali, Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **54**, 5.6.1–5.6.37 (2016). doi: 10.1002/cpbi.3; pmid: 27322406
91. T. I. Croll, ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D Struct. Biol.* **74**, 519–530 (2018). doi: 10.1107/S2059798318002425; pmid: 29872003
92. M. J. Abraham *et al*., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015). doi: 10.1016/j.softx.2015.06.001
93. K. J. Boyd, E. R. May, BUMPy: A model-independent tool for constructing lipid bilayers of varying curvature and composition. *J. Chem. Theory Comput.* **14**, 6642–6652 (2018). doi: 10.1021/acs.jctc.8b00765; pmid: 30431272
94. T. A. Wassenaar, H. I. Ingólfsson, R. A. Böckmann, D. P. Tieleman, S. J. Marrink, Computational Lipidomics with *insane*: A Versatile tool for generating custom membranes for molecular simulations. *J. Chem. Theory Comput.* **11**, 2144–2155 (2015). doi: 10.1021/acs.jctc.5b00209; pmid: 26574417
95. M. Vögele, J. Köfinger, G. Hummer, Molecular dynamics simulations of carbon nanotube porins in lipid bilayers. *Faraday Discuss.* **209**, 341–358 (2018). doi: 10.1039/C8FD00011E; pmid: 29974904
96. R. M. Bhaskara, S. M. Linker, M. Vögele, J. Köfinger, G. Hummer, Carbon nanotubes mediate fusion of lipid vesicles. *ACS Nano* **11**, 1273–1280 (2017). doi: 10.1021/acsnano.6b05434; pmid: 28103440
97. W. Kabsch, C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983). doi: 10.1002/bip.360221211; pmid: 6667333
98. Z. Benayad, S. von Bülow, L. S. Stelzl, G. Hummer, Simulation of FUS protein condensates with an adapted coarse-grained model. *J. Chem. Theory Comput.* **17**, 525–537 (2021). doi: 10.1021/acs.jctc.0c01064; pmid: 33307683
99. M. Javanainen, H. Martinez-Seara, I. Vattulainen, Excessive aggregation of membrane proteins in the Martini model. *PLOS ONE* **12**, e0187936 (2017). doi: 10.1371/journal.pone.0187936; pmid: 29131844
100. L. Monticelli *et al*., The MARTINI coarse-grained force field: Extension to Proteins. *J. Chem. Theory Comput.* **4**, 819–834 (2008). doi: 10.1021/ct700324x; pmid: 26621095
101. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984). doi: 10.1063/1.448118
102. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007). doi: 10.1063/1.2408420; pmid: 17212484
103. M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981). doi: 10.1063/1.328693
104. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, 27–28 (1996). doi: 10.1016/0263-7855(96)00018-5; pmid: 8744570
105. N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, O. Beckstein, MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011). doi: 10.1002/jcc.21787; pmid: 21500218

ACKNOWLEDGMENTS

We acknowledge support from the Electron Microscopy Core Facility (EMCF) and IT services of European Molecular Biology Laboratory (EMBL) Heidelberg. We thank S. Welsch at the Central Electron Microscopy Facility of the Max Planck Institute of Biophysics for technical expertise. We thank T. Hoffman and R. Alves for help with the AlphaFold installation. **Funding:** M.B. acknowledges funding by EMBL, the Max Planck Society, and the European Research Council (ComplexAssembly 724349). J.K. acknowledges funding from the Federal Ministry of Education and Research of Germany (FKZ 031L0100). The work by M.S. and G.H. on computer simulations was supported by the Max Planck Society. M.S. was supported by the EMBL Interdisciplinary Postdoc Programme under Marie Curie COFUND actions. M.S. and G.H. were supported by the Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz (LOEWE) DynaMem program of the State of Hessen. **Author contributions:** S.M. prepared the samples and collected and analyzed the HeLa envelope data. S.M., W.J.H.H., and E.M. collected the HeLa envelope and HeLa in cellulo data. S.M. and B.T. performed the cryo-ET analysis. C.E.Z. prepared, collected, and analyzed the HEK control and NUP210Δ data. R.T. determined the structure of human NUP155. M.-T.M. performed the BioID experiments. F.H.S. and K.B. prepared the HEK NUP210Δ cell line. A.O.-K. performed modeling and prepared figures. M.S. performed the MD simulations. G.H. performed the membrane elastic theory analysis. J.K. and A.O.-K. prepared software. S.M., G.H., J.K., and M.B.

conceptualized the study, supervised the project, and wrote the manuscript. CRediT roles: Conceptualization: S.M., A.O.-K., G.H., J.K., and M.B. Methodology: S.M., A.O.-K., B.T., M.S., and J.K. Investigation: S.M., A.O.-K., M.S., R.T., B.T., C.E.Z., K.B., F.H.S., E.M., M.-T.M., W.H., G.H., J.K., and M.B. Visualization: S.M., A.O.-K., M.S., R.T., C.E.Z., M.-T.M., and J.K. Funding acquisition: M.B. Project administration: G.H., J.K., and M.B. Supervision: S.M., G.H., J.K., and M.B. Writing – original draft: S.M., M.S., G.H., M.B., and J.K. Writing – review & editing: S.M., A.O.-K., M.S., G.H., M.B., and J.K. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** EM maps associated with the manuscript have been deposited in the Electron Microscopy Data Bank (EMDB) under accession codes EMD-14243, -14321, -14322, -14325, -14326, -14327, -14328, and -14330. The models have been deposited into the Protein Data Bank (PDB) under accession codes 7R1Y, 7R5K, and 7R5J. The predicted local quality scores (pLDDT) have been reported in the B-factor property of the deposited models. Software code for subtomogram averaging and geometric cleaning is available at Zenodo (*71*). **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.abm9506
Supplementary Text
Figs. S1 to S29
Tables S1 to S6
References (*106*–*129*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 24 October 2021; accepted 22 April 2022
10.1126/science.abm9506

RESEARCH ARTICLE SUMMARY

NUCLEAR PORE COMPLEX

Architecture of the linker-scaffold in the nuclear pore

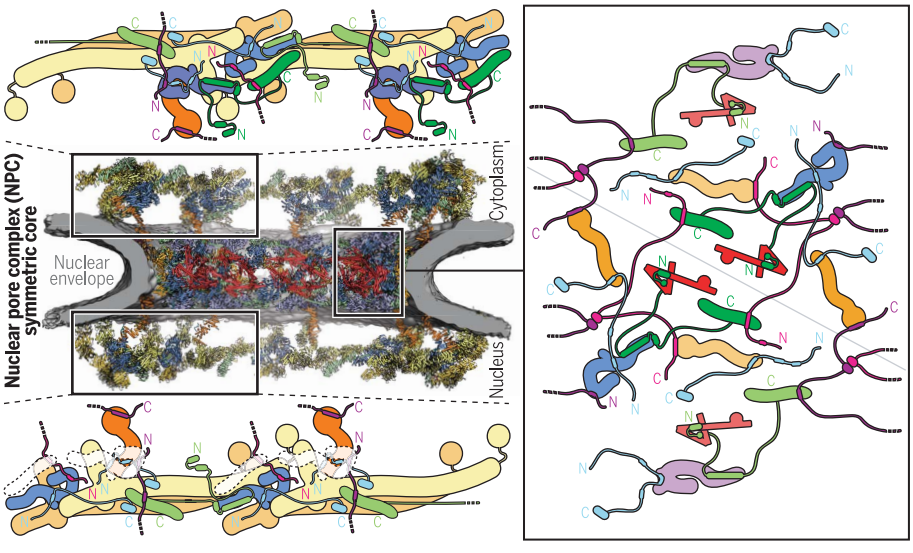
Stefan Petrovic, Dipanjan Samanta†, Thibaud Perriches†, Christopher J. Bley, Karsten Thierbach, Bonnie Brown, Si Nie, George W. Mobbs, Taylor A. Stevens, Xiaoyu Liu, Giovani Pinton Tomaleri, Lucas Schaus, André Hoelz*

INTRODUCTION: In eukaryotic cells, the selective bidirectional transport of macromolecules between the nucleus and cytoplasm occurs through the nuclear pore complex (NPC). Embedded in nuclear envelope pores, the ~110-MDa human NPC is an ~1200-Å-wide and ~750-Å-tall assembly of ~1000 proteins, collectively termed nucleoporins. Because of the NPC's eightfold rotational symmetry along the nucleocytoplasmic axis, each of the ~34 different nucleoporins occurs in multiples of eight. Architecturally, the NPC's symmetric core is composed of an inner ring encircling the central transport channel and two outer rings anchored on both sides of the nuclear envelope. Because of its central role in the flow of genetic information from DNA to RNA to protein, the NPC is commonly targeted in viral infections and its nucleoporin constituents are associated with a plethora of diseases.

RATIONALE: Although the arrangement of most scaffold nucleoporins in the NPC's symmetric core was determined by quantitative docking of crystal structures into cryo–electron tomographic (cryo-ET) maps of intact NPCs, the topology and molecular details of their cohesion by multivalent linker nucleoporins have remained elusive. Recently, in situ cryo-ET reconstructions of NPCs from various species have indicated that the NPC's inner ring is capable of reversible constriction and dilation in response to variations in nuclear envelope membrane tension, thereby modulating the diameter of the central transport channel by ~200 Å. We combined biochemical reconstitution, high-resolution crystal and single-particle cryo–electron microscopy (cryo-EM) structure determination, docking into cryo-ET maps, and physiological validation to elucidate the molecular architecture of the linker-scaffold interaction network that not only is essential for the NPC's integrity but also confers the plasticity and robustness necessary to allow and withstand such large-scale conformational changes.

RESULTS: By biochemically mapping scaffold-binding regions of all fungal and human linker nucleoporins and determining crystal and single-particle cryo-EM structures of linker-scaffold complexes, we completed the characterization of the biochemically tractable linker-scaffold network and established its evolutionary conservation, despite considerable sequence divergence. We determined a series of crystal and single-particle cryo-EM structures of the intact Nup188 and Nup192 scaffold hubs bound to their Nic96, Nup145N, and Nup53 linker nucleoporin binding regions, revealing that both proteins form distinct question mark–shaped keystones of two evolutionarily conserved hetero-octameric inner ring complexes. Linkers bind to scaffold surface pockets through short defined motifs, with flanking regions commonly forming additional disperse interactions that reinforce the binding. Using a structure-guided functional analysis in *Saccharomyces cerevisiae*, we confirmed the robustness of linker-scaffold interactions and established the physiological relevance of our biochemical and structural findings. The near-atomic composite structures resulting from quantitative docking of experimental structures into human and *S. cerevisiae* cryo-ET maps of constricted and dilated NPCs structurally disambiguated the positioning of the Nup188 and Nup192 hubs in the intact fungal and human NPC and revealed the topology of the linker-scaffold network. The linker-scaffold gives rise to eight relatively rigid inner ring spokes that are flexibly interconnected to allow for the formation of lateral channels. Unexpectedly, we uncovered that linker-scaffold interactions play an opposing role in the outer rings by forming tight cross-link staples between the eight nuclear and cytoplasmic outer ring spokes, thereby limiting the dilatory movements to the inner ring.

CONCLUSION: We have substantially advanced the structural and biochemical characterization of the symmetric core of the *S. cerevisiae* and human NPCs and determined near-atomic composite structures. The composite structures uncover the molecular mechanism by which the evolutionarily conserved linker-scaffold establishes the NPC's integrity while simultaneously allowing for the observed plasticity of the central transport channel. The composite structures are roadmaps for the mechanistic dissection of NPC assembly and disassembly, the etiology of NPC-associated diseases, the role of NPC dilation in nucleocytoplasmic transport of soluble and integral membrane protein cargos, and the anchoring of asymmetric nucleoporins. ■



Linker-scaffold architecture in the human NPC's symmetric core. Near-atomic composite structure of the NPC's symmetric core obtained by quantitative docking of high-resolution crystal and single-particle cryo-EM structures into a cryo-ET reconstruction of the intact human NPC. Schematic representations of the intricate linker-scaffold topology of the cytoplasmic outer ring, inner ring, and nuclear outer ring (clockwise from top) are depicted for the boxed regions. C, C terminus; N, N terminus.

Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA.
***Corresponding author. Email: hoelz@caltech.edu**
†These authors contributed equally to this work.
Cite this article as S. Petrovic *et al.*, *Science* 376, eabm9798 (2022). DOI: 10.1126/science.abm9798

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abm9798

RESEARCH ARTICLE

NUCLEAR PORE COMPLEX

Architecture of the linker-scaffold in the nuclear pore

Stefan Petrovic, Dipanjan Samanta†, Thibaud Perriches†‡, Christopher J. Bley, Karsten Thierbach§, Bonnie Brown, Si Nie, George W. Mobbs, Taylor A. Stevens, Xiaoyu Liu¶, Giovani Pinton Tomaleri, Lucas Schaus, André Hoelz*

Nuclear pore complexes (NPCs) mediate the nucleocytoplasmic transport of macromolecules. Although the arrangement of the structured scaffold nucleoporins in the NPC's symmetric core has been determined, their cohesion by multivalent unstructured linker nucleoporins has remained elusive. Combining biochemical reconstitution, high-resolution structure determination, docking into cryo–electron tomographic reconstructions, and physiological validation, we elucidated the architecture of the evolutionarily conserved linker-scaffold, yielding a near-atomic composite structure of the human NPC's ~64-megadalton symmetric core. Whereas linkers generally play a rigidifying role, the linker-scaffold of the NPC provides the plasticity and robustness necessary for the reversible constriction and dilation of its central transport channel and the emergence of lateral channels. Our results substantially advance the structural characterization of the NPC symmetric core, providing a basis for future functional studies.

The enclosure of genetic material in the nucleus requires the selective transport of folded proteins and ribonucleic acids across the nuclear envelope, for which the nuclear pore complex (NPC) is the sole gateway (*1*–*4*). Beyond its function as a selective, bidirectional channel for macromolecules, the role of the NPC extends to genome organization, transcription regulation, mRNA maturation, and ribosome assembly (*1*, *2*). The NPC and its components are implicated in the etiology of many human diseases, including viral infections (*5*, *6*). The building blocks of the NPC are a set of ~34 different proteins collectively termed nucleoporins (nups). In the NPC, nups assemble into defined subcomplexes that are generally present in multiples of eight, adding up to a mass of ~110 MDa in the human NPC (*1*–*4*). The NPC architecture consists of a symmetric core with asymmetric decorations on its nuclear and cytoplasmic faces (Fig. 1A). The symmetric core displays eight- and twofold rotational symmetry about the nucleocytoplasmic axis and axes coplanar with the nuclear envelope, respectively. It consists of two outer rings that sit on top of the nuclear envelope and an inner ring that lines the lumen generated by the fusion of the two lipid bilayers of the nuclear envelope. From the inner ring, unstructured phenylalanine-glycine (FG) repeats are projected into the central transport channel to establish the diffusion barrier. Although ~40 kDa has historically been considered the threshold for passive diffusion (*7*, *8*), the size selectivity of the barrier shows a more gradual dependence on molecular mass (*9*). Active transport is generally mediated by karyopherins, whose affinity for FG repeats and ultrafast exchange kinetics allow karyopherin-bound cargo to traverse the diffusion barrier (*10*–*13*).

The structural characterization of the NPC has progressed through efforts to reconstitute and crystallize ever larger portions of it, from small nup domain fragments to complexes as large as the ~400-kDa heteroheptameric Y-shaped coat nup complex (CNC) (*14*–*30*). In parallel, progress has been driven by efforts to push the resolution of cryo–electron tomographic (cryo-ET) reconstructions of intact NPCs (*31*). The docking of the CNC crystal structure into an ~32-Å cryo-ET map of the intact human NPC was the first demonstration that biochemical reconstitution and crystal structures could be used to interpret cryo-ET maps and unraveled the head-to-tail tandem arrangement of CNCs in the outer rings (*29*, *32*). The reconstitution and piecemeal structural analysis of two heterononameric ~425-kDa inner ring complexes provided the basis for docking 17 symmetric core nups into an ~23-Å cryo-ET map of the intact human NPC (*28*, *33*), yielding a near-atomic composite structure of the entire ~56-MDa symmetric core of the human NPC (*34*, *35*). Subsequently, a similar approach elucidated the near-atomic architectures of constricted and dilated states of the *Saccharomyces cerevisiae* NPC using crystal structures to interpret ~25-Å cryo-ET maps (*36*, *37*). Apart from an additional distal CNC ring and associated nups present in the outer rings of the human NPC, the human and *S. cerevisiae* NPCs present equivalent nup arrangements (*34*–*38*).

The inner ring of the human NPC is composed of six scaffold nups called NUP155, NUP188, NUP205, NUP54, NUP58, and NUP62; two primarily unstructured linker nups called NUP53 and NUP98; and NUP93, which is a hybrid of both (*34*, *35*). The doughnut-shaped inner ring adopts a concentric cylinder architecture, in which membrane-anchored NUP155 forms the outermost coat, followed by layers of NUP93, NUP205 or NUP188, and the NUP54•NUP58•NUP62 channel nucleoporin heterotrimer (CNT) in the center, providing the FG repeats to form the diffusion barrier in the central transport channel. Unlike the extensive interactions of large, folded domains found in the CNC (*16*–*20*, *22*, *25*, *27*, *29*, *39*), the structured domains of the inner ring nups do not interact directly. Instead, the inner ring is held together by the linker nups NUP53 and NUP98 and the linker region of NUP93, which are proposed to connect the scaffolds of the four layers (*28*, *33*, *35*, *40*–*42*). The resulting linker-scaffold architecture allows for a substantial ~200-Å dilation of the inner ring's central transport channel, accompanied by the generation of lateral channels between the eight spokes, as observed in recent cryo-ET analyses of purified and in situ human and fungal NPCs (*36*, *43*–*46*). The linker-scaffold is expected to play a fundamental role in establishing an architectural framework to accommodate the structural changes associated with the reversible constriction and dilation of the inner ring.

Whereas our previous work achieved the identification of most of the scaffold nup locations in the NPC, a comprehensive understanding and the molecular details of the linker-scaffold interaction network that mediates the cohesion of the symmetric core have remained elusive. Here, we report the characterization of the complete tractable set of linker-scaffold interactions through residue-level biochemical mapping of scaffold-binding regions in linker nups and the determination of crystal and single-particle cryo–electron microscopy (cryo-EM) structures of linker-scaffold complexes. Our analysis revealed a common linker-scaffold binding mode, whereby linkers are anchored by central structured motifs whose binding is reinforced by disperse interactions of flanking regions. We quantitatively docked the complete set of linker-scaffold structures into an ~12-Å cryo-ET reconstruction of the human NPC (provided by Martin Beck's group) (*47*) and an ~25-Å in situ cryo-ET reconstruction of the *S. cerevisiae* NPC (*36*). In the inner ring, our new linker-scaffold structures allowed for the unambiguous assignment of NUP188 and NUP205 to 16 peripheral and 16 equatorial positions, respectively. From the nuclear envelope to the central transport

Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA.
***Corresponding author. Email: hoelz@caltech.edu**
†These authors contributed equally to this work.
‡Present address: Care Partners, 146 chemin de l'Etang, 69380 Dommartin, France.
§Present address: Odyssey Therapeutics, Inc., Industriepark Höchst, G875, 65926 Frankfurt am Main, Germany.
¶Present address: Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, 609 Charles E. Young Drive East, Los Angeles, CA 90095, USA.

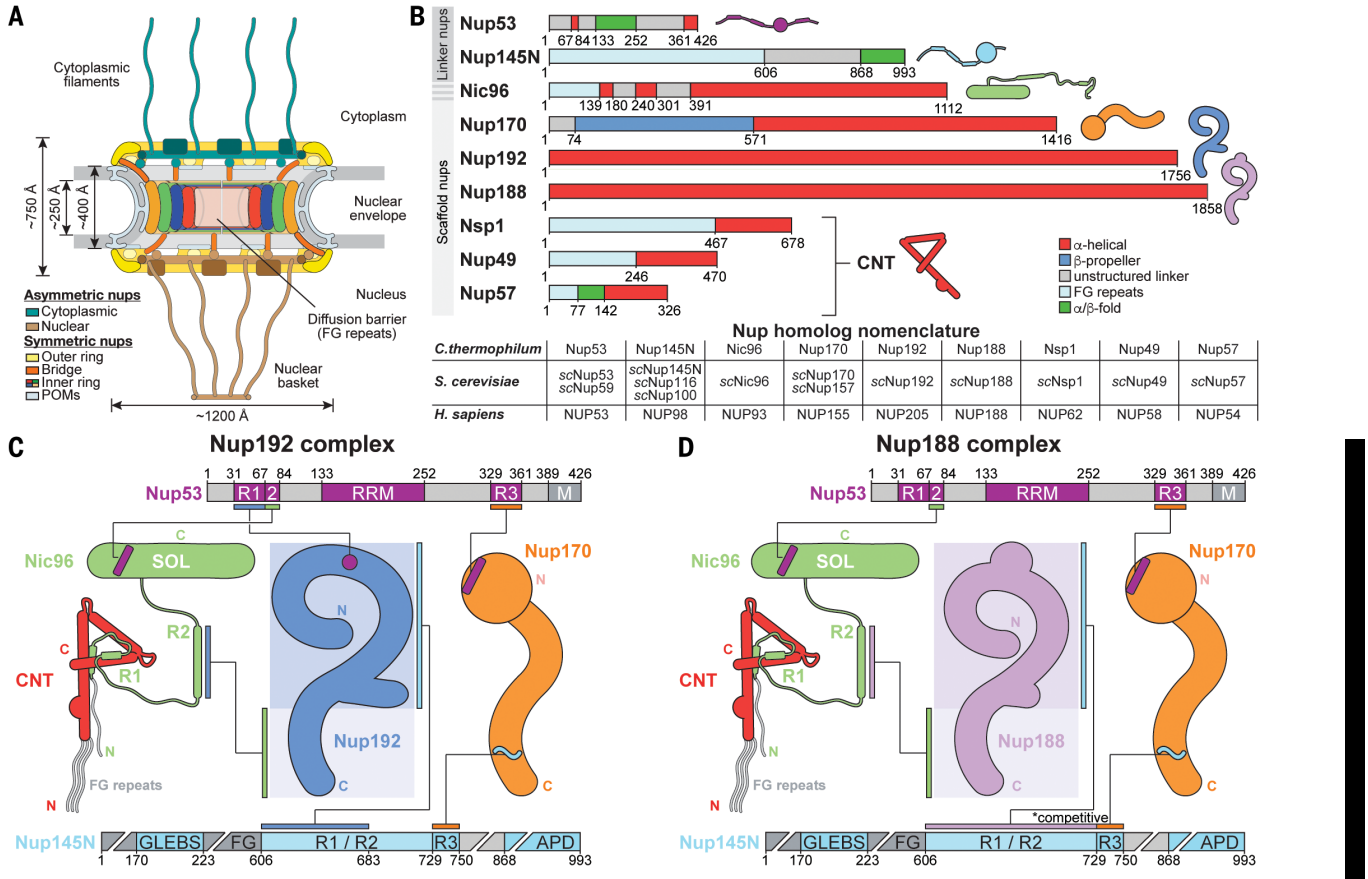


Fig. 1. Outline of the symmetric core inner ring architecture. (**A**) Cross-sectional schematic of the NPC architecture. POMs, integral membrane proteins of the pore membrane domain. (**B**) Domain structures of the *C. thermophilum* inner ring linker and scaffold nups. Nic96 consists of linker (residues 1 to 390) and scaffold (residues 391 to 1112) regions. Nomenclature for nup homologs from *C. thermophilum*, *S. cerevisiae*, and *H. sapiens* is indicated. Multiple paralogs exist for some *S. cerevisiae* nups. (**C** and **D**) Schematic map of previously established linker-scaffold interactions in alternative, mutually exclusive inner ring complexes organized around Nup192 and Nup188 scaffold hubs (*35*). Black lines connecting colored bars indicate interactions between nup regions. C, C terminus; N, N terminus.

channel, linkers bridge the layers of the inner ring to coalesce scaffold nups into eight relatively rigid spokes that are flexibly interconnected, allowing for the formation of lateral channels. The linker-scaffold confers the plasticity necessary for the reversible dilation and constriction of the inner ring in response to alterations in nuclear envelope membrane tension. The topology of linker-scaffold interactions between inner ring nups is conserved from fungi to humans. We carried out systematic functional analyses of the linker-scaffold network, including the development of a minimal linker *S. cerevisiae* strain, establishing its robustness and essential nature. Our quantitative docking analysis of the human NPC revealed eight NUP205•NUP93 complexes that cross-link adjacent spokes in both nuclear and cytoplasmic outer rings. Facing the central transport channel, additional eight NUP205•NUP93 copies are exclusively anchored at the base of the cytoplasmic outer ring. NUP93 emerges as a versatile linker-scaffold hybrid that recruits and positions the FG repeat-harboring CNT to the inner ring and reinforces the tandem head-to-tail CNC arrangement in the outer rings, explaining its fundamental role in maintaining the integrity of the entire NPC. Our analysis substantially advances the structural characterization of the ~64-MDa symmetric core and lays out a roadmap for future studies on the NPC assembly and function.

Results

Biochemical and structural analysis of Chaetomium thermophilum linker-scaffold interactions

Nup192 and Nup188 are question mark–shaped scaffold keystones of two alternative eight-protein inner ring complexes, both of which include the linkers Nup145N, Nup53, and the scaffold-linker hybrid Nic96 (Fig. 1, B to D) (*28*, *33*, *35*, *40*, *41*). Previously, a composite structure of the full-length ~200-kDa Nup192 was determined by superposing overlapping structures of its N- and C-terminal parts, revealing an extended α-helical solenoid with a question mark–shaped architecture, composed of 5 HEAT and 15 ARM repeats and a prominent central Tower (*35*, *40*, *41*, *48*). High-resolution structures of Nup188 encompassing the ~130-kDa N-terminal domain (NTD), which contains a central SH3-like domain insertion, and the ~45-kDa C-terminal Tail region again revealed extended α-helical solenoids composed of ARM and HEAT repeats (*28*, *49*). However, no structural information could so far be obtained for the ~32-kDa Nup188 central region equivalent to the Nup192 Tower. Although binding of Nup192 and Nup188 has been biochemically mapped to the Nic96, Nup53, and Nup145N linkers, the molecular details are unknown. To gather these details, which are crucial to the elucidation of the linker-scaffold architecture, we performed the following comprehensive biochemical and structural analyses.

Nup192 interaction with Nic96

Using a crystallizable Nup192$^{\Delta Head}$ (residues 153 to 1756) fragment (*35*), we obtained cocrystals with our previously biochemically mapped Nic96$^{187–301}$ fragment that diffracted to 3.6-Å resolution (fig. S1). Nic96 residues 187 to 239 were not resolved and were found to be dispensable for Nup192 binding by isothermal titration calorimetry (ITC), because both Nic96$^{187–301}$ and Nic96^{R2} (residues 240 to 301) retained a dissociation constant (K_D) of ~75 nM (Fig. 2E

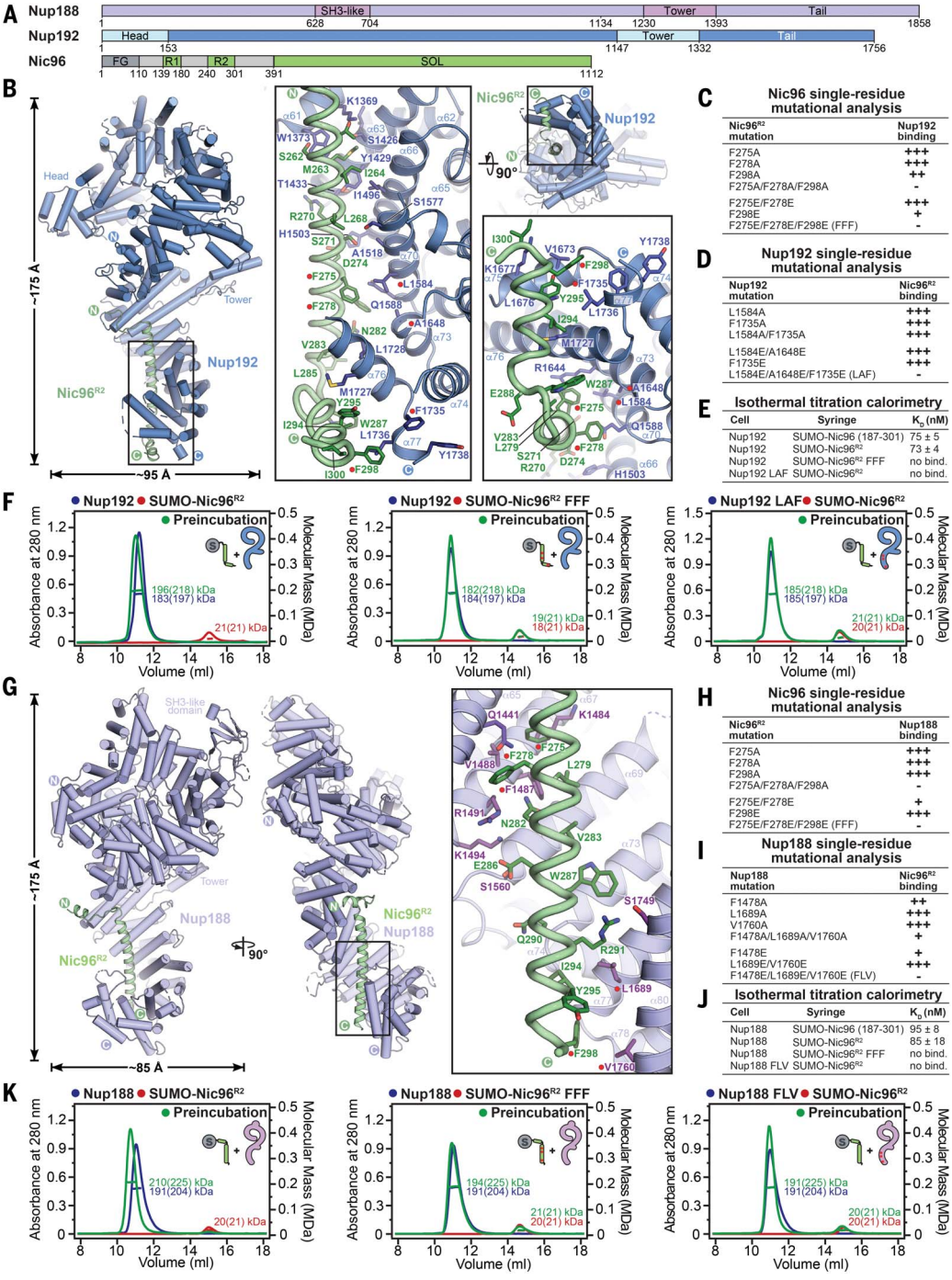


Fig. 2. Structural and biochemical analyses of the Nup192-Nic96 and Nup188-Nic96 interactions. (**A**) Domain structures of *C. thermophilum* Nup188, Nup192, and Nic96. (**B**) Cartoon representation of the 3.8-Å *C. thermophilum* Nup192•Nic96^{R2} single-particle cryo-EM structure. Inset regions are magnified to illustrate the molecular details of the Nup192-Nic96 interaction. Red circles indicate residues involved in the Nup192-Nic96 interaction. (**C** and **D**) Summary of the effect of structure-guided (C) SUMO-Nic96^{R2} and (D) Nup192 mutations on Nup192•SUMO-Nic96^{R2} complex formation, assayed by SEC. +++, no effect; ++, weak effect; +, moderate effect; –, abolished binding. (**E**) K_Ds determined by triplicate ITC experiments, with the mean and associated standard error reported. (**F**) SEC-MALS interaction analyses of Nup192•SUMO-Nic96^{R2} and interaction-abolishing mutants. Measured molecular masses are indicated, with theoretical masses in parentheses. S, small ubiquitin-like modifier (SUMO). (**G**) Two views in cartoon representation of the 4.4-Å *C. thermophilum* Nup188•Nic96^{R2} crystal structure phased and built with the 2.8-Å Nup188^{NTD} and 3.4-Å Nup188^{Tail} [Protein Data Bank (PDB) ID 5CWU] (*28*) crystal structures. Inset regions are magnified to illustrate the molecular details of the Nup188-Nic96^{R2} interaction. Red circles indicate residues involved in Nup188-Nic96^{R2} binding. (**H** and **I**) Summary of the effect of structure-guided (H) SUMO-Nic96^{R2} and (I) Nup188 mutations on Nup188•SUMO-Nic96^{R2} complex formation, assayed by SEC. (**J**) K_Ds determined by triplicate ITC experiments with the mean and associated standard error reported. (**K**) SEC-MALS analysis of Nup188•SUMO-Nic96^{R2} and interaction-abolishing mutants. 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.

and fig. S2, A and D). The Nic96^{R2} sequence register was unambiguously assigned by identifying seleno-L-methionine (SeMet)–labeled residues in anomalous difference Fourier maps (fig. S1). To obtain the structure of full-length Nup192•Nic96^{R2}, we determined a single-particle cryo-EM reconstruction at 3.8-Å global resolution from a refined set of 176,609 particles whose preferential orientation resulted in an anisotropic 3.5- to 4.0-Å directional Fourier shell correlation (FSC) resolution range (Fig. 2B and fig. S3). Nic96^{R2} forms two amphipathic α helices connected by a sharply kinking loop that extend from the midpoint to the base of the Nup192 question mark–shaped α-helical solenoid. The longer N-terminal α helix is cradled by the concave Nup192 surface formed by 10 ARM and HEAT repeats and the central Tower, whereas the shorter C-terminal α helix packs against a hydrophobic patch formed from the C-terminal Nup192 α helices α75 to α77 (ARM-20). Comparison of our crystal and cryo-EM structures confirmed the molecular details of Nic96^{R2} binding and identified a conformational difference in the width of the gap between the Nup192 Head and Tower subdomains (fig. S4). To validate the molecular details of the Nup192-Nic96^{R2} interface, we performed structure-guided mutagenesis and assessed binding by size exclusion chromatography coupled to multiangle light scattering (SEC-MALS) and ITC. Consistent with an ~3700-Å^2 hydrophobic interface, binding was not strongly affected by individual substitutions and was only abolished by Nic96 FFF or Nup192 LAF combination mutations (F, Phe; L, Leu; A, Ala; Fig. 2, C to F, and figs. S2 and S5 to S7).

Nup188 interaction with Nic96

Next, we tested whether the same Nic96 region was sufficient for Nup188 binding. Indeed, ITC measurements revealed that Nup188 binds both Nic96^{R2} and Nic96$^{187–301}$ with similar K_Ds of ~90 nM (Fig. 2J and fig. S8, A and D). We determined crystal structures of Nup188•Nic96^{R2} and Nup188^{NTD} (residues 1 to 1134) at 4.4- and 2.8-Å resolution, respectively, the latter of which aided with phasing and model building. The Nup188 and Nic96^{R2} sequence registers were unambiguously assigned by identifying SeMet-labeled residues in anomalous difference Fourier maps (Fig. 2G and fig. S9). Like Nup192, Nup188 adopts an overall question mark–shaped architecture, composed of an N-terminal Head subdomain, 9 HEAT repeats, 13 ARM repeats, and a central, comparatively more compact Tower (Fig. 2G). Similarly, Nic96^{R2} binds a concave surface between the midpoint and the base of the Nup188 question mark–shaped α-helical solenoid, burying ~3700 Å^2 of combined surface area. Although Nup188-bound Nic96^{R2} also forms two amphipathic α helices, the α helices start and end at different residues, resulting in a secondary structure that radically differs from the Nup192-bound form (Fig. 2, B and G). Nup188-bound Nic96^{R2} has a shorter N-terminal helix that binds to the central Tower and a longer C-terminal helix cradled in the concave surface at the base of Nup188 (Fig. 2G). Notably, the Nic96^{R2} FFF mutation that abolishes Nup192 binding had the same effect on Nup188 binding, despite the structural polymorphism between Nup192- and Nup188-bound Nic96^{R2} (Fig. 2, H to K, and figs. S8B, S10, and S12). Analogous to the Nup192 LAF mutant, we identified a triple Nup188 FLV substitution that disrupted Nic96^{R2} binding (V, Val; Fig. 2, I to K, and figs. S8C, S11, and S12).

Nup192 interaction with Nup145N and Nup53

To identify the Nup145N regions necessary and sufficient for Nup192 binding, we performed a five-alanine scanning mutagenesis and truncation analysis of Nup145N (Fig. 3A). Substituting five consecutive residues at a time to alanines, we found a hotspot between residues 626 and 655 that displayed diminished binding to Nup192 (Fig. 3A and fig. S13). N- and C-terminal Nup145N truncation resulted in a minimal Nup145N^{R1} peptide (residues 616 to 683) that recapitulated the Nup192-Nup145N interaction, although shorter Nup145N fragments showed residual binding to Nup192 (Fig. 3B and fig. S14). ITC measurements confirmed that Nup192 binding is primarily sustained by Nup145N's R1 region, with K_Ds of ~825 and ~1600 nM for Nup145N and Nup145N^{R1} binding, respectively (Fig. 3G and fig. S15).

With our previously mapped minimal Nup53^{R1} fragment (residues 31 to 67) (*41*), we reconstituted an ~220-kDa Nup192•Nic96^{R2}•Nup145N^{R1}•Nup53^{R1} complex and obtained a single-particle cryo-EM reconstruction at 3.2-Å global resolution from a selected set of 484,910 particles whose preferential orientation resulted in an anisotropic 3.1- to 3.6-Å directional FSC resolution range (Fig. 3C and fig. S16). For Nup53^{R1}, the cryo-EM map only resolved the central phenylalanine-glycine (FG) dipeptide buried in a hydrophobic pocket at the top of the Nup192 molecule. Key contacts involve Leu^{441} and Trp^{499} of Nup192 and Phe^{48} of Nup53, consistent with our previous identification of these residues as required for the Nup53^{R1}-Nup192 interaction by systematic mutagenesis (Fig. 3D) (*41*). The Nup145N^{R1} binding site is proximal to the Nic96^{R2} binding site, at the midpoint of the question mark–shaped Nup192, where a hydrophobic pocket anchors the Nup145N^{R1} MYKL motif (residues 633 to 636; M, Met; Y, Tyr; K, Lys) that runs perpendicular to the long axis of the question mark, with its N terminus oriented toward the N terminus of Nic96^{R2} (Fig. 3E). Overall, comparison of the Nup192 structures in complex with different linkers demonstrated that linker binding does not induce conformational rearrangements in the scaffold Nup192 (fig. S17).

Validation of the Nup192-Nup145N^{R1} interface through structure-guided mutagenesis confirmed the importance of the central hydrophobic Nup145N MYKL anchor motif (Fig. 3F and fig. S18), but complete ablation of binding was only observed when the three flanking basic residues on either side were also mutated to alanine in the 10-residue KKR-MYKL-RKR mutant (R, Arg) (Fig. 3, F to H, and figs. S15 and S18 to S20). Conversely, mutagenesis of the Nup145N^{R1} MYKL binding site in Nup192 identified a quadruple Nup192 LIFH mutant that specifically abolished Nup192 binding to Nup145N^{R1} but not Nic96^{R2} or Nup53 (I, Ile; H, His; Fig. 3, F and H, and figs. S19 to S21). Although basic residues flanking both Nup145N's MYKL and Nup53's FG (*41*) anchor motifs contribute to Nup192 binding, flanking residues were not resolved in the cryo-EM density.

Nup188 interaction with Nup145N

To identify the Nup145N regions necessary and sufficient for Nup188 binding, we used a five-alanine scanning and fragment truncation approach analogous to the mapping of the Nup145N-Nup192 interaction. This identified a minimal Nup145N^{R2} peptide (residues 640 to 732) that recapitulated wild-type binding and a region between residues 706 and 715 that affected Nup188^{NTD} binding upon five-alanine substitution (Fig. 4, A and B, and figs. S22 and S23). Consistent with our previous findings (*28*), we also confirmed that Nup188 does not bind Nup53, even in the presence of Nic96^{R2} and Nup145N (fig. S24).

We determined the structure of an ~220-kDa Nup188•Nic96^{R2}•Nup145N^{R2} complex by single-particle cryo-EM. An initial set of 709,123 preferentially oriented particles produced a reconstruction of Nup188•Nic96^{R2} at 2.4-Å global resolution and an anisotropic 2.3- to 2.5-Å directional FSC resolution range. Local three-dimensional classification of particles based on emergent excess density at the top of the question mark–shaped Nup188 molecule identified a subset of 298,317 particles that yielded a reconstruction of Nup188•Nic96^{R2}•Nup145N^{R2} at 2.8-Å global resolution and an anisotropic 2.7- to 2.9-Å directional FSC resolution range (Fig. 4C and fig. S25). Nup145N^{R2} buries residues Ile^{709}, Leu^{710}, and Phe^{715} in a hydrophobic cradle adjacent to the SH3-like domain. As with Nup145N^{R1} bound to Nup192, only a central portion of the Nup145N^{R2} peptide was resolved (residues 706 to 718) (Fig. 4C). No meaningful conformational changes were observed between the different Nup188 structures, in response to linker binding (fig. S26).

Substitution of two resolved Nup145N residues, Leu^{710}→Ala (L710A) and Phe^{715}→Ala

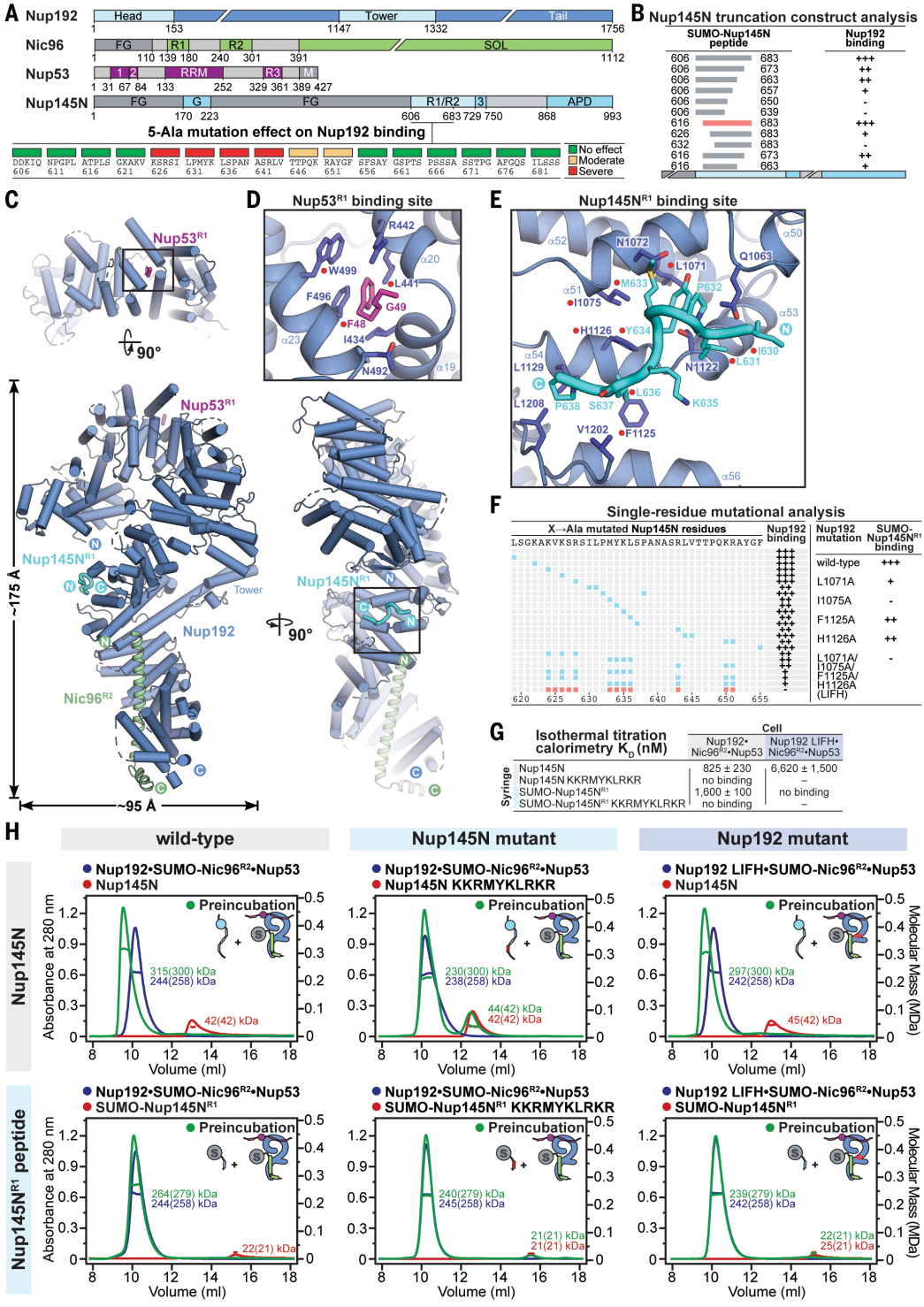


Fig. 3. Structural and biochemical analyses of the Nup192-Nup145N interaction. (**A**) Domain structures of *C. thermophilum* Nup192, Nic96, Nup53, and Nup145N and the effect of each five-alanine substitution on Nup145N binding to Nup192, as assessed by SEC and indicated by colored boxes above the Nup145N primary sequence. (**B**) Summary of SEC binding analysis identifying the minimal Nup145N^{R1} (red) region sufficient for Nup192 binding. +++, no effect; ++, weak effect; +, moderate effect; –, abolished binding. (**C** to **E**) Cartoon representation of (C) the 3.2-Å *C. thermophilum* Nup192•Nic96^{R2}•Nup53^{R1}•Nup145N^{R1} single-particle cryo-EM structure. Insets indicate regions magnified to illustrate molecular details of (D) the Nup192-Nup53^{R1} and (E) the Nup192-Nup145N^{R1} interactions. Red circles indicate residues involved in the Nup192-Nup53^{R1} (*41*) and Nup192-Nup145N^{R1} interactions. (**F**) Effect of Nup145N alanine substitutions (cyan squares) on Nup192 binding, assayed by SEC (left). Effect of structure-guided Nup192 alanine substitutions on SUMO-Nup145N^{R1} binding, assayed by SEC (right). (**G**) K_Ds determined by triplicate ITC experiments, with the mean and associated standard error reported. (**H**) SEC-MALS analysis of Nup192•SUMO-Nic96^{R2}•Nup53•Nup145N and Nup192•SUMO-Nic96^{R2}•Nup53•SUMO-Nup145N^{R1} complex formation and disruption by mutants. Measured molecular masses are indicated, with respective theoretical masses provided in parentheses.

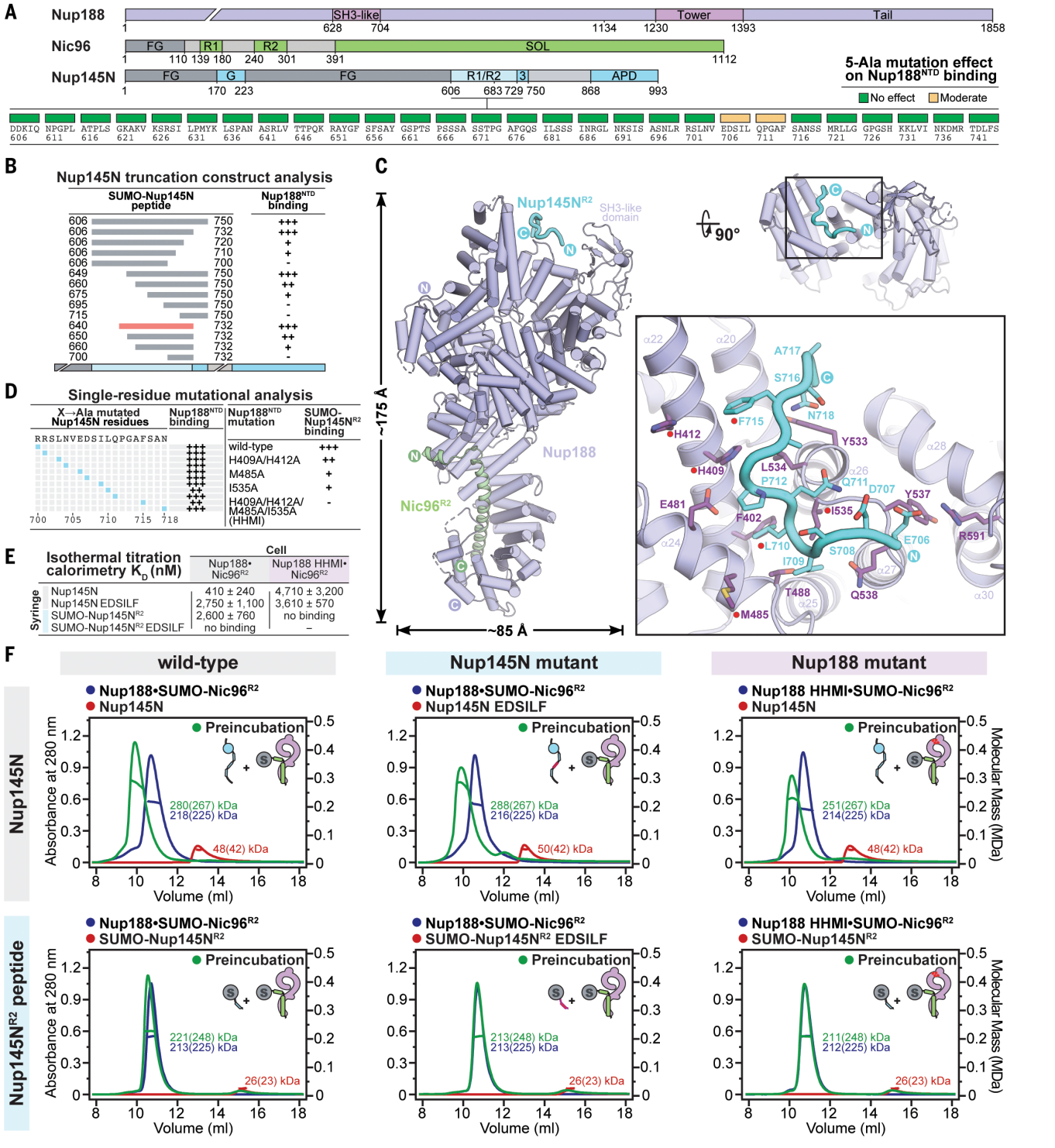


Fig. 4. Structural and biochemical analyses of the Nup188-Nup145N interaction. (**A**) Domain structures of *C. thermophilum* Nup188, Nic96, and Nup145N and the effect of each five-alanine substitution on Nup145N binding to Nup188^{NTD}, as assessed by SEC and indicated by colored boxes above the Nup145N primary sequence. (**B**) Summary of SEC binding analysis identifying the minimal Nup145N^{R2} (red) region sufficient for binding to Nup188^{NTD}. +++, no effect; ++, weak effect; +, moderate effect; –, abolished binding. (**C**) Cartoon representation of the 2.8-Å *C. thermophilum* Nup188•Nic96^{R2}•Nup145N^{R2} single-particle cryo-EM structure. Inset indicates region magnified to illustrate molecular details of the Nup188-Nup145N^{R2} interaction. Red circles indicate residues involved in the Nup188-Nup145N^{R2} interaction. (**D**) Effect of Nup145N alanine substitutions (cyan squares) on Nup188^{NTD} binding, assayed by SEC (left). Effect of structure-guided Nup188^{NTD} alanine substitutions on SUMO-Nup145N^{R2} binding, assayed by SEC (right). (**E**) K_Ds determined by triplicate ITC experiments, with the mean and associated standard error reported. (**F**) SEC-MALS analysis of Nup188•SUMO-Nic96^{R2}•Nup145N and Nup188•SUMO-Nic96^{R2}•SUMO-Nup145N^{R2} complex formation and disruption by mutants. Measured molecular masses are indicated, with respective theoretical masses provided in parentheses.

(F715A), moderately disrupted Nup188^{NTD} binding (Fig. 4D and fig. S27). Further systematic mutagenesis led to a Nup145N EDSILF mutant, which respectively abolished and reduced Nup188 binding to Nup145N^{R2} and Nup145N (E, Glu; D, Asp; S, Ser; Fig. 4, E and F, and figs. S28 to S30). Structure-guided mutagenesis of Nup188 residues interfacing with Nup145N^{R2} identified a Nup188 HHMI mutant that abolished binding to Nup145N^{R2} but not to Nup145N (Fig. 4, D to F, and figs. S28 to S30). Overall, the greater tolerance of the Nup188-Nup145N interaction to binding site mutations demonstrates an even greater reliance on promiscuous binding events in flanking regions dispersed well beyond the structurally resolved core anchor motif.

Comparison of the Nup192- and Nup188-linker complexes

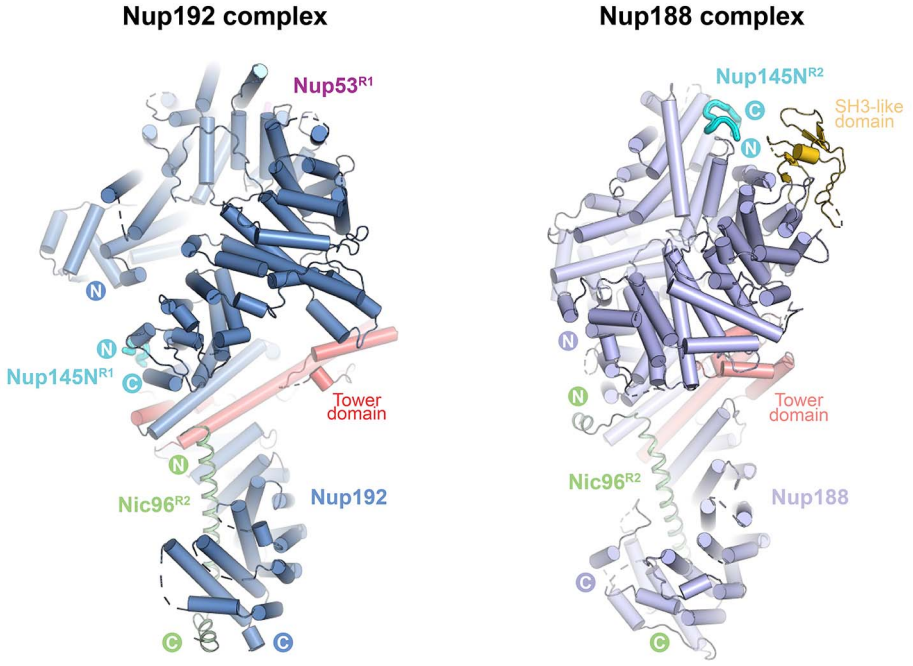
The determination of full-length structures of both Nup192 and Nup188 scaffolds bound to their respective linkers permits a direct comparison of these two distantly homologous α-helical solenoids (~28% sequence similarity) (Movie 1). Although both structures share the same overall question mark–shaped architecture, the Nup188 α-helical solenoid displays a tighter superhelical twist, resulting in an ~10-Å narrower molecule with a compacted N-terminal ring (fig. S31, A and B). A Tower protrudes from the midpoint of the α-helical solenoid toward the Head subdomain in both structures, extending further in Nup192 than the comparatively compressed Nup188 version. Nic96^{R2} binds both scaffolds at the base of the question mark but notably adopts different secondary structures, switching between which requires breaking and reforming of α helices (fig. S31C). By contrast, Nup145N binds to different parts of Nup192 and Nup188, at the midpoint and the top of the question mark–shaped molecules, respectively. Interestingly, the Nup145N^{R2} binding site at the top of Nup188 is nearly congruent with that of the Nup53^{R1} on Nup192 (fig. S31A).

Our structures and biochemical analysis identify two distinct types of linker-scaffold interactions. Nic96^{R2} binds with high affinity, using the same well-defined ~60-residue motif in binding to both Nup192 and Nup188. On the contrary, Nup145N binds to Nup192 and Nup188 through protracted, overlapping binding regions and a distinctive common binding mode: Both interactions depend on a structurally defined ~10-residue Nup145N anchor motif, yet tight binding requires extensive ~20- to 60-residue N- and C-terminal flanking regions with high basic character. The Nup53-Nup192 interaction relies on a similar binding mode. The evasiveness of these interaction-enhancing linker flanking regions to structural characterization suggests that their binding to scaffold surfaces is highly dynamic and promiscuous. Notably, the previously characterized ~14-residue Nup170-binding motif of Nup145N (residues 729 to 750) does not depend on binding-enhancing flanking regions, suggesting that the uncovered mode of Nup145N and Nup53 binding to the Nup192 and Nup188 scaffolds is a desirable evolutionary outcome and architectural principle of the NPC inner ring.

Together, these data complete the structural and biochemical characterization of the biochemically tractable linker-scaffold interactions. Nup145N binds at distinct sites on Nup192 and Nup188, forming mutually exclusive interactions with either Nup192 and Nup170 or Nup188 through extensive overlapping binding sites mapped to Nup192 (R1, residues 616 to 683), Nup188 (R2, residues 640 to 732) and Nup170 (R3, residues 729 to 750). Binding by means of a central anchor motif enhanced by extensive flanking regions is reminiscent of Velcro, in which weak binding events accumulate to build a robust yet flexible interaction with manifold productive binding configurations possible in terms of both spatial distribution and occupancy. As an architectural principle, Velcro-like binding could accommodate scaffold movements without entirely breaking the linker-scaffold, maintaining the NPC's integrity in face of large-scale dilation or constriction.

Architecture of the S. cerevisiae linker-scaffold

The inner ring of the NPC contains eightfold rotational symmetry about a nucleocytoplasmic axis and twofold symmetry in the plane of the nuclear envelope (*34*, *35*). In the *S. cerevisiae* (*sc*) NPC, each of the 16 inner ring protomers were proposed to consist of a *sc*Nup192 and a *sc*Nup188 inner ring complex (Fig. 1, C and D), with *sc*Nup192 and *sc*Nup188 located at the equatorial and peripheral positions, respectively (*36*, *37*). High-confidence quantitative docking of our full-length experimental Nup192•Nic96^{R2}•Nup145N^{R1}•Nup53^{R1} and Nup188•Nic96^{R2}•Nup145N^{R2} structures in an ~25-Å in situ cryo-ET map of the *S. cerevisiae* NPC (fig. S32) (*36*) confirmed these proposals. Whereas docking of the folded scaffolds Nup170, Nic96, Nup192, Nup188, and the CNT into cryo-ET maps of intact NPCs revealed their positioning to form four concentric cylinders, the linker network that connects them has remained elusive. Combined with our previously determined structures of Nup170•Nup53^{R3}, Nup170•Nup145N^{R3}, Nic96•Nup53^{R2}, and CNT•Nic96^{R1} (*28*, *35*), the Nup192•Nic96^{R2}•Nup145N^{R1}•Nup53^{R1} and Nup188•Nic96^{R2}•Nup145N^{R2} structures allowed us to identify the locations of all scaffold-bound linker regions. We considered whether the length of linker polypeptides connecting pairs of scaffold-bound linker segments constrained the topology of linker connections. We found a single topology connecting linker segments related by the shortest Euclidean distance. For a detailed description of these



Movie 1. Structural analysis of the Nup192 and Nup188 inner ring complexes. Comparison of crystal and single-particle cryo-EM structures of *C. thermophilum* Nup192 and Nup188 scaffolds in complex with Nic96, Nup145N, and Nup53 linkers. Cryo-EM densities are rendered as isosurfaces colored according to their assigned protein chain.

results, see the supplementary text (figs. S32 to S37).

Together, these data elucidated the architecture of the *S. cerevisiae* inner ring linker-scaffold (Fig. 5A and Movie 2). Apart from spoke bridging mediated by Nup53 orthologs, all linker-scaffold connections in the *S. cerevisiae* inner ring occur within the same spoke, thereby allowing interspoke gaps to form. The linker-scaffold architecture provides a molecular explanation for the inner ring's ability to exist in constricted and dilated states (*36*, *37*).

The S. cerevisiae linker-scaffold is robust and essential

Owing to ancestral gene duplication events in *S. cerevisiae*, there are several linker and scaffold nup paralogs, including *sc*Nup145N paralogs *sc*Nup116 and *sc*Nup100, *sc*Nup53 paralog *sc*Nup59, and *sc*Nup170 paralog *sc*Nup157 (Fig. 1B) (*2*). The *sc*Nup100 and *sc*Nup116 paralogs contain sequences homologous to the Nup192, Nup188, and Nup170 binding regions characterized in *C. thermophilum* Nup145N, but only *sc*Nup116 possesses the Gle2 binding site (GLEBS) motif (fig. S38) (*50*).

To interrogate the function of individual scaffold-binding regions in the linker *sc*Nup116, we established a *S. cerevisiae* minimal *nup100Δnup116Δnup145Δ* strain complemented with *sc*Nup116 and *sc*Nup145C, ectopically expressed from centromeric plasmids (Fig. 5, B and C, and fig. S39). Next, we systematically mutated all functional elements in the *sc*Nup116 sequence, including the *sc*Nup192, *sc*Nup188, and *sc*Nup157/170 scaffold-binding regions R1, R2, and R3, respectively, with three types of mutations: deletions (Δ*R1*, Δ*R2*, and Δ*R3*), substitutions with glycine-serine (GS) linkers of equivalent length (*R1/40×GS*, *R2/40×GS*, and *R3/12×GS*), or substitutions of sequence-conserved residues shown to disrupt binding of the *C. thermophilum* Nup145N to the respective scaffolds (*R1m*, *R2m*, and *R3m*) (Fig. 5B and fig. S38). Deletions and GS-linker substitutions, being aggressive types of mutations, were lethal if targeting R1 and affected growth, mRNA export, and 60*S* preribosome export if introduced in R2 and R3. The less aggressive combination of substitutions, *R1m*, caused substantial yet non-lethal phenotypic effects, which were further exacerbated through combination with *R2m* (*R1m+R2m*) or *R3m* (*R1m+R3m*), culminating with the lethal *R1m+R2m+R3m* triple mutation (Fig. 5, D and E, and fig. S40). Interestingly, all *sc*Nup116 mutations resulted in temperature-dependent loss of enhanced green fluorescent protein (eGFP)–*sc*Nup116 from the nuclear envelope rim and concomitant emergence of eGFP-*sc*Nup116 foci (Fig. 5D and fig. S40F), as previously reported (*51*–*53*).

Next, we transposed the insight from our structural and biochemical characterization of the *C. thermophilum* linker-scaffold into equivalent substitutions of conserved residues or more aggressive truncations of binding site–harboring subdomains of *sc*Nup192, *sc*Nup188, and *sc*Nic96 (Fig. 5F). Notably, the combination of *LAF* and *LIFH* substitutions (*LAF+LIFH*) that ablated Nup192 binding to Nic96^{R2} and Nup145N^{R1}, respectively, failed to rescue the lethal *nup192Δ* phenotype (Fig. 5G and fig. S41 and S42). Analogously, the combination of *FLV* and *HHMI* substitutions (*FLV+HHMI*) that ablated Nup188 binding to Nic96^{R2} and Nup145N^{R2}, respectively, led to an additive cold-sensitive slow-growth phenotype with mRNA and 60*S* preribosome export defects in the synthetic lethal *nup188Δpom34Δ* strain (Fig. 5, G and H, and figs. S43 and S44) (*54*–*56*). Finally, we introduced the transposed *FFF* substitutions of evolutionarily conserved hydrophobic residues that abolished Nic96^{R2} binding to Nup192 and Nup188 into *sc*Nic96, along with *sc*Nic96^{R2} deletion (Δ*R2*) (Fig. 5F and fig. S45). Surprisingly, neither mutation resulted in a detectable phenotype in a *nic96Δ* strain (Fig. 5, F to H, and fig. S46). The composite structure of the NPC linker-scaffold suggests that Nic96^{R2} binding to Nup192 and Nup188 restricts the diffusive path of the N-terminal Nic96 linker, thereby correctly positioning the Nic96^{R1} assembly sensor that recruits the CNT complex (Fig. 5J). We reasoned that CNT mispositioning would affect the spatial distribution and local concentration of FG repeats, with consequences on nucleocytoplasmic transport. Therefore, we replaced the Nic96^{R2} region with GS-linkers matching the number of residues (*R2/66×GS*) or approximating its α-helical length (*R2/32×GS*) (Fig. 5F). Despite not affecting CNT recruitment by the Nic96^{R1} region, the *R2/66×GS* and *R2/32×GS* mutations resulted respectively, in lethal and severely deleterious effects on growth, mRNA export, and 60*S* preribosome export (Fig. 5, G to I, and fig. S46). For a detailed description of these results, see the supplementary text.

Taken together, these data demonstrate the physiological relevance of our residue-level biochemical and structural characterization of Nup192 and Nup188 as keystone scaffold hubs of the inner ring that integrate connections between the membrane-coating Nup170 layer and the central transport channel–interfacing CNT layer through respective interactions with Nup145N and the N-terminal Nic96 linker regions. The wild-type phenotype of the *nup53Δnup59Δ* strain (*57*) precludes analysis of the *sc*Nup53 and *sc*Nup59 interactions in *S. cerevisiae*. However, this fact, coupled with our knockout of all but one of the *sc*Nup145N paralogs, highlights the robustness of the *S. cerevisiae* inner ring architecture, which can tolerate a considerable loss of linker-scaffold interactions. Robustness is also found in nup-nup interactions, whereby perturbing a linker-scaffold interaction requires multiple residue substitutions in both structured motifs and flanking linker regions.

Evolutionary conservation of the human linker-scaffold

Despite low sequence conservation, composite structures of the human and *S. cerevisiae* NPCs reveal an identical positioning of the scaffold nups, suggesting that the linker-scaffold architecture is evolutionarily conserved (*34*–*37*). Specifically, the human linker-scaffold interactions, the topology of scaffold-binding regions in the linkers, and the location of linker-binding sites in the scaffolds are expected to match those of the *C. thermophilum* nups (*28*, *33*, *35*, *40*, *41*). We developed expression and purification protocols for recombinant human nups (Fig. 6A), enabling systematic interaction analyses between scaffold and linker nups, for which we generated truncation and sequence variants, aided by multispecies sequence alignments (figs. S38, S45, and S47 to S49). For a detailed description of these results, see the supplementary text (figs. S50 to S53). Together with our previous mapping of the NUP155^{CTD}-NUP98^{R3} interaction (*35*), these data establish that the linker-scaffold is evolutionarily conserved from *C. thermophilum* to humans, including the linker-binding sites in the scaffolds and the topology of the scaffold-binding regions in the linkers (Fig. 6B).

Biochemical and structural analysis of the human NUP93-NUP53 interaction

Our dissection of the human linker-scaffold interaction network identified an interaction between NUP93^{SOL} and a NUP53 region N-terminal of the RNA recognition motif (RRM)–like domain (N) (residues 1 to 169) (fig. S51) that was nevertheless devoid of homology to the corresponding *C. thermophilum* Nup53^{R2} amphipathic α-helix motif that fits into a hydrophobic groove of the Nic96^{SOL} scaffold (figs. S48 and S49) (*35*). Through fragment truncation and five-alanine scanning mutagenesis, we identified a NUP53^{R2} region (residues 84 to 150) that formed a stable complex with NUP93^{SOL}, within which residues 86 to 100 were required for binding to NUP93^{SOL} (Fig. 6, C and D, and figs. S54 and S55).

To elucidate the molecular details of binding between the divergent NUP53^{R2} and NUP93^{SOL}, we determined crystal structures of apo NUP93^{SOL} and NUP93^{SOL}•NUP53^{R2} at 2.0- and 3.4-Å resolution, respectively. As with other linker-scaffold interactions, only a core region (residues 88 to 95) of the biochemically mapped minimal NUP53^{R2} was resolved (Fig. 6E and fig. S56). *C. thermophilum* and human Nic96^{SOL} orthologs display equivalent α-helical solenoid architectures (Fig. 6F) (*35*, *58*, *59*). In contrast to the *C. thermophilum* Nup53^{R2}

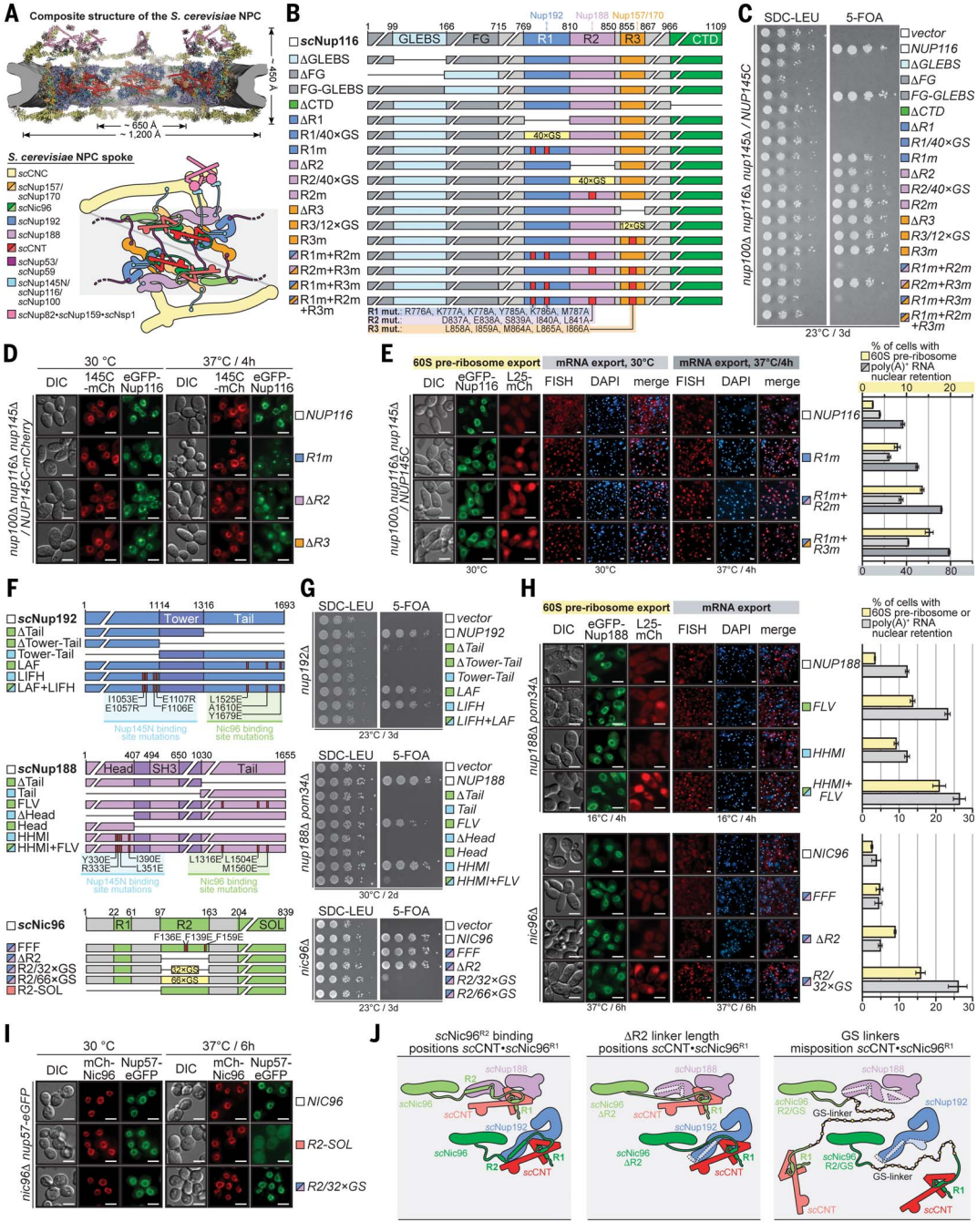
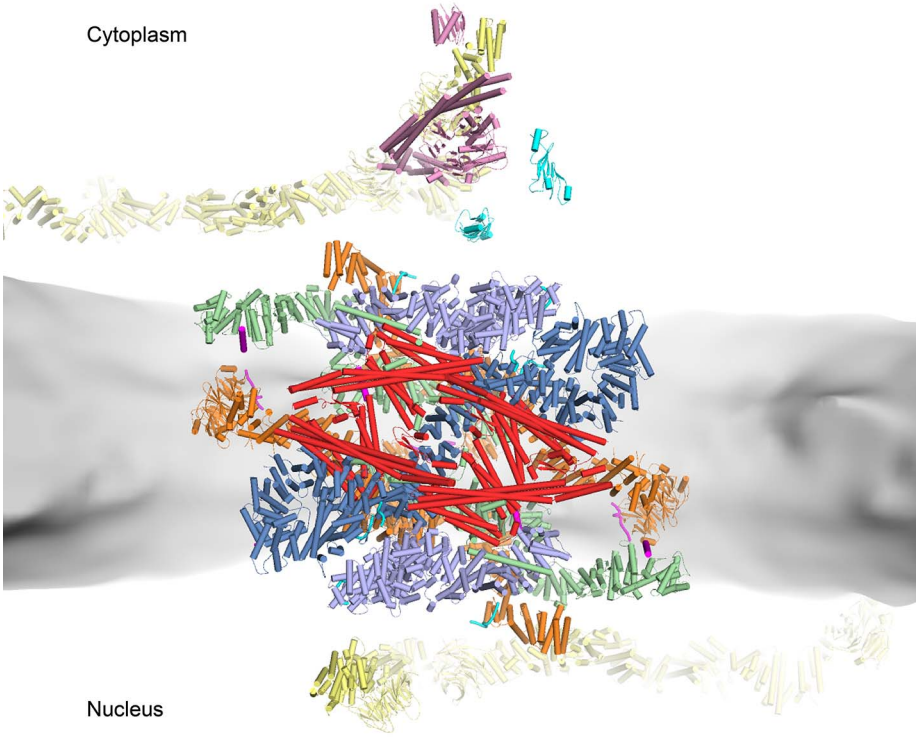


Fig. 5. Functional in vivo dissection of the *S. cerevisiae* NPC linker-scaffold. (**A**) Cross-sectional view of the *S. cerevisiae* NPC composite structure generated by docking linker-scaffold structures into an ~25-Å in situ subtomogram averaged cryo-ET map [Electron Microscopy Data Bank (EMDB) ID EMD-10198] (*36*) (top). Schematic representation of a *S. cerevisiae* NPC spoke (bottom). (**B**) Domain structure of *sc*Nup116 variants, the Gle2-binding sequence (GLEBS), FG repeats, the *sc*Nup192-binding region (R1), the *sc*Nup188-binding region (R2), and the *sc*Nup157/170-binding region (R3). (**C**) Viability analysis of a 10-fold dilution series of a *nup100Δnup116Δnup145Δ/NUP145C* strain expressing *sc*Nup116 variants and subjected to 5-fluoroorotic acid (5-FOA) selection for loss of rescuing wild-type plasmid. (**D**) Subcellular localization at permissive (30°C) and growth-challenging (37°C) temperatures of a representative subset of eGFP-*sc*Nup116 variants in a *nup100Δnup116Δnup145Δ/NUP145C-mCherry* strain. (**E**) Representative images and quantitation (n > 500) of subcellular localization of 60*S* preribosomal export reporter *sc*Rpl25-mCherry and poly(A)$^+$ RNA at 30° and 37°C in the presence of a representative subset of *sc*Nup116 variants. (**F**) Domain structures of *sc*Nup192, *sc*Nup188, and *sc*Nic96 variants. (**G**) Viability analysis of a 10-fold dilution series of *nup192Δ*, *nup188Δpom34Δ*, and *nic96Δ S. cerevisiae* strains expressing *sc*Nup192, *sc*Nup188, and *sc*Nic96 variants, respectively, and subjected to 5-FOA selection for loss of rescuing wild-type plasmids. (**H**) Representative images and quantitation (n > 500) of the nuclear *sc*Rpl25-mCherry and poly(A)$^+$ RNA retention in the presence of *sc*Nup188 and *sc*Nic96 variants at indicated growth-challenging temperatures in *nup188Δpom34Δ* and *nic96Δ S. cerevisiae* strains, respectively. (**I**) Subcellular localization at permissive (30°C) and growth-challenging (37°C) temperatures of the *sc*CNT subunit *sc*Nup57-eGFP in the presence of mCherry-*sc*Nic96 variants. (**J**) Schematic model of CNT positioning in wild-type, *sc*Nic96^{R2} deletion, and GS-linker replacement strains. Squares associated with variant labels are color coded according to the nup binding partners targeted by the mutation. All experiments were performed in triplicate. Mean and associated standard error are reported for all quantitation. Scale bars are 5 μm.



Movie 2. Architecture of the *S. cerevisiae* NPC linker-scaffold. An animated dissection of the composite structure generated by docking high-resolution crystal and single-particle cryo-EM structures into the *S. cerevisiae* ~25-Å NPC cryo-ET map (EMDB ID EMD-10198) (*36*). The nuclear envelope and protein cryo-ET densities are rendered as opaque and transparent gray isosurfaces, respectively. Crystal structures of nups and nup complexes are shown in cartoon representation. Unstructured linker connections between docked scaffolds are drawn as dashed lines.

amphipathic α helix, human NUP53^{R2} binds the conserved NUP93^{SOL} hydrophobic groove that encompasses α helices α5 and α13 to α15 as a linear eight-residue motif, burying ~1100 Å^2 of combined surface area (Fig. 6, E and I to K, Movie 3, and fig. S56E). Systematic alanine substitution of the resolved NUP53^{R2} motif, invariant across metazoan NUP53 sequences (fig. S49), confirmed the key role of the Pro^{89}-Pro^{90} di-proline and Ile^{94}, consequently also illustrating the importance of the NUP93^{SOL} residues that interface with them (Fig. 6, G, H, and L, and fig. S57). Together, our data establish that despite distinct binding modes and low sequence conservation, linker-scaffold interactions are evolutionarily conserved, further highlighting their essential role and indicating that shape conservation of scaffolds is a key determinant of the NPC architecture.

Architecture of the human NPC symmetric core

Quantitative docking of nup complexes in cryo-ET maps of the human NPC

We have previously demonstrated that nup ortholog crystal structures can be successfully used to interpret the density of an ~23-Å cryo-ET map of the intact human NPC, yielding a near-atomic composite structure of the NPC symmetric core that included linker-scaffold crystal structures of the Nup170•Nup53^{R3}•Nup145N^{R3}, Nic96^{SOL}•Nup53^{R2}, and CNT•Nic96^{R1} complexes (*28*, *35*). The newly available structures of full-length Nup192 and Nup188 as part of Nup192•Nic96^{R2}•Nup145N^{R1}•Nup53^{R1} and Nup188•Nic96^{R2}•Nup145N^{R2} linker-scaffold complexes, as well as the human NUP93^{SOL}•NUP53^{R2} complex, allowed us to build on our previous analysis with an improved ~12-Å cryo-ET map of the intact human NPC (provided by Martin Beck's group) (*47*). As for the *S. cerevisiae* NPC described above, our quantitative docking approach consisted of statistically scoring the fit of resolution-matched densities simulated from crystal and single-particle cryo-EM structures that were randomly placed and locally refined in cryo-ET maps of the human NPC. Structures of the CNC, Nup192•Nic96^{R2}•Nup145N^{R1}•Nup53^{R1}, Nup188•Nic96^{R2}•Nup145N^{R2}, and NUP358^{NTD} (reported in the accompanying manuscript) (*60*) were readily placed in cryo-ET maps of the entire NPC or of the inner ring portion. Assigned density was then iteratively subtracted from the maps to reduce the subsequent search space for NUP93^{SOL}•NUP53^{R2}, Nup170•Nup53^{R3}•Nup145N^{R3}, CNT•Nic96^{R1}, and NUP53^{RRM} (fig. S58). For a detailed description of these results, see the supplementary text (figs. S58 to S77).

The structures and improved cryo-ET map of the intact human NPC have disambiguated the placement of NUP188 and NUP205 hubs in the inner ring and distal outer ring positions, led to the discovery of a proximal NUP205 in the cytoplasmic outer ring, identified NUP93^{SOL} in the outer rings, placed NUP53^{RRM} homodimers between inner ring spokes, and revealed a comprehensive map of scaffold-bound linker segments that implied a single symmetric core linker topology connecting linker segments related by the shortest Euclidean distance (Movie 4 and figs. S77 and S78). The composite structure includes ~400,000 ordered residues that explain nearly all protein density of the symmetric core and assign the protein identity and location of ~64 MDa out of ~110 MDa of the human NPC mass.

Architecture of cytoplasmic and nuclear outer rings

In both the cytoplasmic and nuclear outer rings of the human NPC, 16 copies of the Y-shaped CNC are arranged in two concentric proximal and distal rings. At equivalent locations on both the cytoplasmic and nuclear sides, eight copies of NUP205 are intercalated between the proximal NUP75 arms and distal NUP107 stalks of CNCs from adjacent spokes. Eight NUP93^{SOL} copies are inserted between the distal NUP107 and proximal NUP96 α-helical solenoids, bisecting the stalks of tandem-arranged CNCs of a single spoke (Fig. 7 and fig. S79). Stretched out, the ~25-residue unstructured linker connecting the R2 and SOL regions of NUP93 bridge the ~95-Å gap between the distal NUP205-bound NUP93^{R2} and the distal NUP93^{SOL} of an adjacent spoke, thus cross-linking the outer ring spokes (Fig. 7, Movie 4, and figs. S78 to S80). Compared to the Nup192 ortholog, NUP205 presents an additional ~240 residues that elongate the C-terminal Tail region, suggesting that ~95 Å is an upper estimate for the distance between distal NUP93^{R2} and NUP93^{SOL} from adjacent spokes.

Specific to the cytoplasmic face, an additional eight copies of the proximal NUP205 are lodged between the NUP75 arm of the proximal CNC and the bridge NUP155 that connects the outer and inner ring (Fig. 7A and fig. S79). The proximal NUP205-bound NUP93^{R2} can only be linked with a proximal NUP93^{SOL} of the same spoke (Fig. 7A and figs. S79 and S80A). Furthermore, the arrangement of NUP53 binding sites on NUP93^{SOL} and NUP205 copies in the cytoplasmic outer rings is compatible with the NUP53-mediated linkage of the distal NUP205 with the proximal NUP93^{SOL} and, conversely, the proximal NUP205 with the distal NUP93^{SOL} (Fig. 7 and

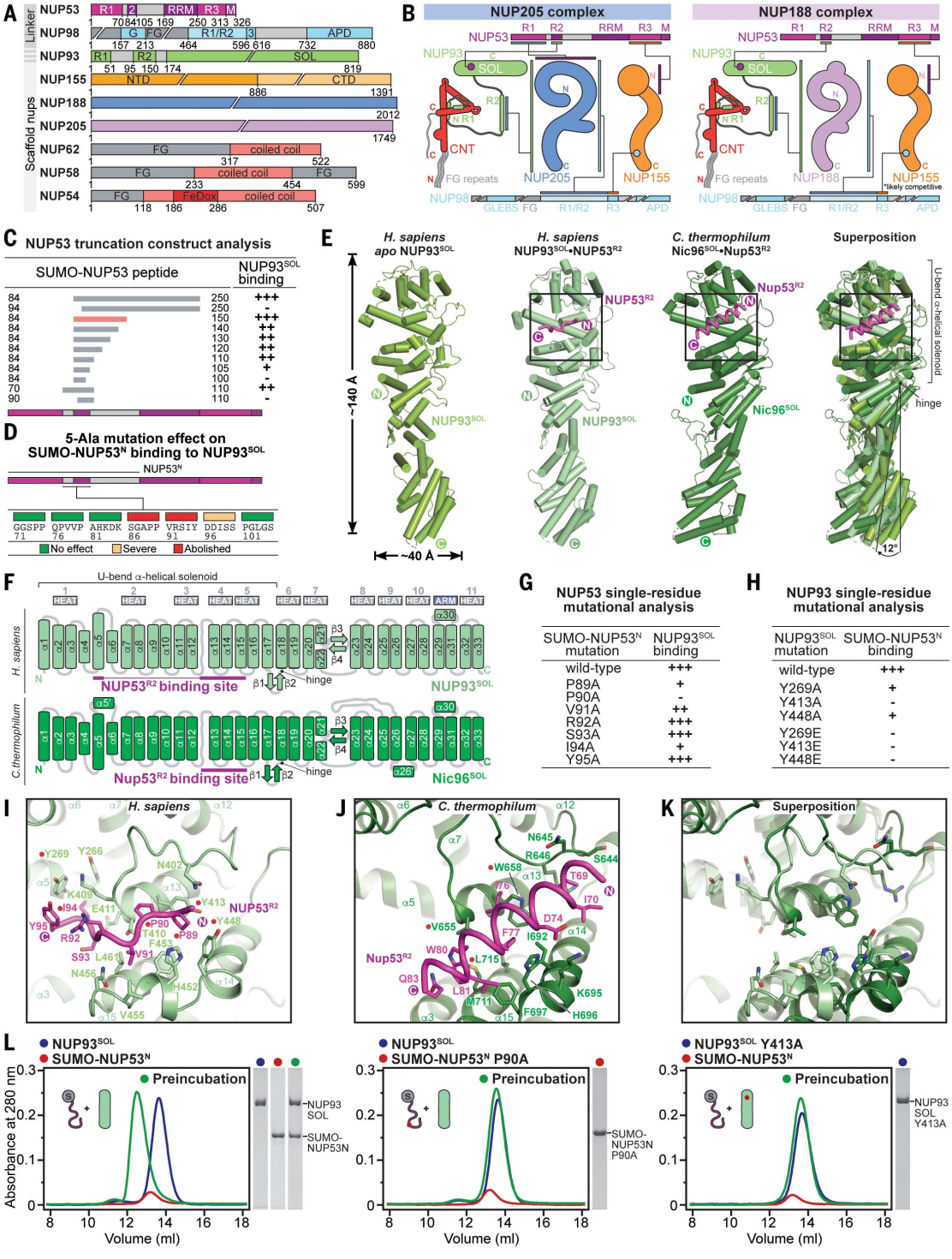
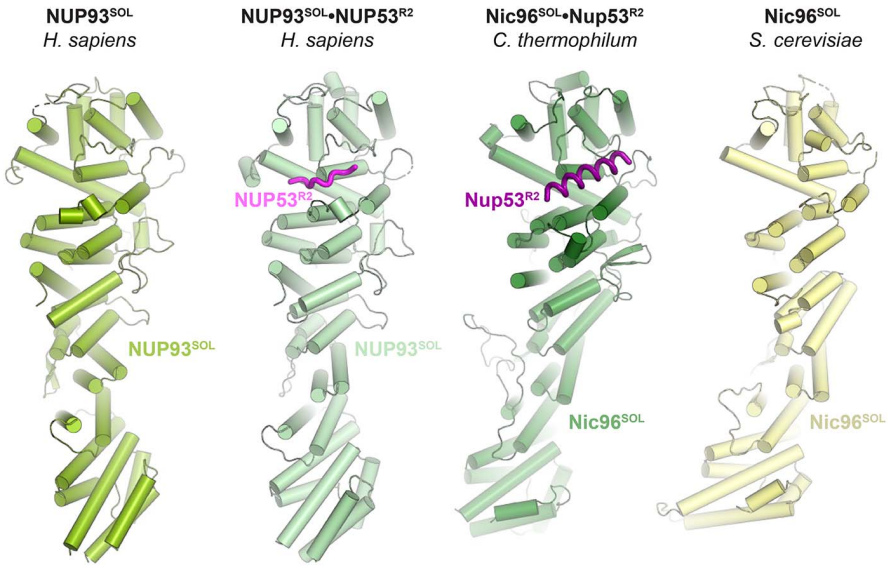
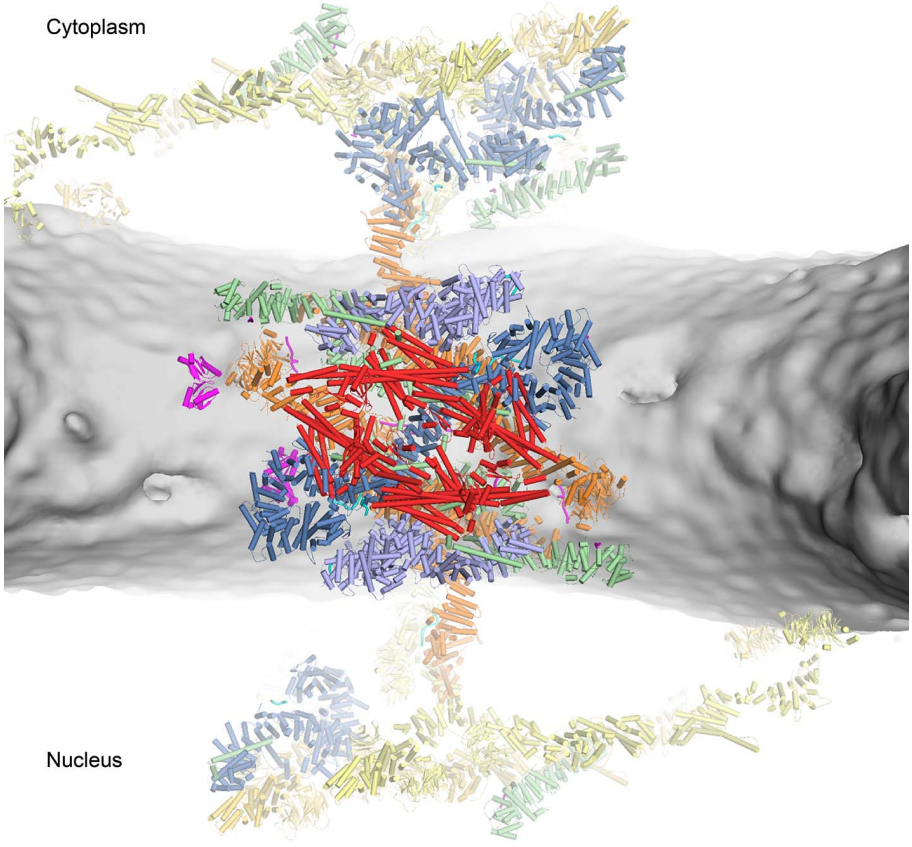


Fig. 6. Evolutionary conservation of the human linker-scaffold network. (**A**) Domain structures of the human inner ring nups. NUP93 consists of linker (residues 1 to 173) and scaffold (residues 174 to 819) regions. (**B**) Schematic summary of the linker-scaffold interactions in complexes organized around the NUP188 and NUP205 scaffold hubs. Black lines connecting colored bars indicate interactions between nup regions. (**C**) Summary of SEC binding analysis identifying the minimal NUP53^{R2} region (red) sufficient for NUP93^{SOL} binding. +++, no effect; ++, weak effect; +, moderate effect; −, abolished binding. (**D**) Effect of each five-alanine substitution on SUMO-NUP53^{N} binding to NUP93^{SOL}, as assessed by SEC and indicated by colored boxes above NUP53 primary sequence. (**E**) Cartoon representations of the 2.0-Å *H. sapiens* apo NUP93^{SOL}, 3.4-Å *H. sapiens* NUP93^{SOL}•NUP53^{R2}, and 2.7-Å *C. thermophilum* Nic96^{SOL}•Nup53^{R2} (PDB ID 5HB3) (*35*) crystal structures and their superposition. An ~12° displacement of the C-terminal α-helical solenoid, pivoted about the hinge loop, is observed between the apo NUP93^{SOL} and NUP93^{SOL}•NUP53^{R2} structures. (**F**) Schematic of the human NUP93^{SOL} and *C. thermophilum* Nic96^{SOL} fold architectures. (**G** and **H**) Summary of the effect of structure-guided mutations in (G) SUMO-NUP53^{N} and (H) NUP93^{SOL} on NUP93^{SOL}•SUMO-NUP53^{N} complex formation, as assayed by SEC. (**I** to **K**) Magnified views of the regions indicated with insets in (E), which compare molecular details of the NUP53^{R2} and Nup53^{R2} binding sites. (**L**) SEC analyses of NUP93^{SOL}•SUMO-NUP53^{N} complex formation and disruption by mutants. SDS–polyacrylamide gel electrophoresis (SDS-PAGE) gel strips of peak fractions visualized by Coomassie staining.



Movie 3. Structural analysis of the NUP93^{SOL}•NUP53^{R2} structure. Comparison of cartoon representations of *H. sapiens* NUP93^{SOL}•NUP53^{R2} with *S. cerevisiae* Nic96^{SOL} (PDB ID 2QX5) (*59*) and *C. thermophilum* Nic96^{SOL}•Nup53^{R2} (PDB ID 5HB3) (*35*) orthologs and comparison of conformational differences between apo NUP93^{SOL} and NUP93^{SOL}•NUP53^{R2} obtained from different crystal forms.



Movie 4. Architecture of the symmetric core of the human NPC. An animated dissection of the composite structure generated by quantitatively docking high-resolution crystal and single-particle cryo-EM structures into the human ~12-Å NPC cryo-ET map (EMDB ID EMD-14322) (*47*). The nuclear envelope and protein cryo-ET densities are rendered as opaque and transparent gray isosurfaces, respectively. Crystal structures of nups and nup complexes are shown in cartoon representation. Unstructured linker connections between docked scaffolds are drawn as dashed lines.

figs. S79 and S80, D and E). NUP53 could also mediate long-range links between the bridge NUP155 and the R1 and R2 binding sites on outer ring NUP205 and NUP93^{SOL}, respectively (fig. S80, F, H, and I).

On both the nuclear and cytoplasmic sides, the NUP98^{APD}-binding NUP96 sites present in the 16 CNC copies recruit 16 copies of NUP98 that can simultaneously satisfy the outer ring NUP205 binding sites because of the ~115-residue linker between the autoproteolytic domain (APD) and the R1-R2-R3 regions (Fig. 7 and fig. S79). The cytoplasmic proximal NUP205 and bridge NUP155 scaffolds could, in principle, be linked by NUP98, although NUP98^{R3} is likely outcompeted from its bridge NUP155 binding site by the asymmetric cytoplasmic filament nups GLE1•NUP42 (*61*), as explained in the accompanying manuscript (Fig. 7A and fig. S80C) (*60*). To maximize nup copy parsimony while satisfying all available scaffold binding sites, the outer rings would recruit 16 copies of NUP53 and NUP98 on each side of the NPC.

Architecture of the inner ring linker-scaffold

The quantitative docking confirmed evolutionary conservation of the inner ring linker-scaffold architecture between *Homo sapiens* and *S. cerevisiae* NPCs (Fig. 8). A NUP155•NUP53 linker-scaffold coats the nuclear envelope, anchored by membrane curvature–sensing amphipathic lipid packing sensor (ALPS) motifs and the C-terminal NUP53 amphipathic helix (*35*, *62*–*64*), with a peripheral and equatorial copy on each side of a spoke midplane. The homodimerizing NUP53^{RRM} domains link spoke halves across the midplane (Fig. 8C and figs. S81 and S82, A and B). A second cross-midplane link between NUP53 and NUP93 connects NUP188•NUP93•CNT and NUP205•NUP93•CNT modules to the NUP155 coat at equatorial and peripheral positions, respectively (Fig. 8, D to F, and fig. S82, C to J). Restrained by their length, NUP93 N-terminal linkers connect NUP188 with the peripheral and NUP205 with the equatorial inner ring NUP93^{SOL} and CNT copies of the same inner ring spoke (Fig. 8, D to F, and fig. S82, G to J).

The equatorial position of NUP205 is further solidified by a NUP98-mediated linkage with the equatorial NUP155 and by a NUP53-mediated linkage with a peripheral NUP93 from an adjacent spoke (Fig. 8E and fig. S82, E and F). As in the *S. cerevisiae* NPC, the NUP205, NUP188, and peripheral NUP155 binding sites for the NUP98 R1, R2, and R3 regions, respectively, are too far apart to be linked by the same NUP98, suggesting that three NUP98 copies are required to satisfy all binding sites on each side of a spoke midplane instead. The NUP98^{APD}-binding NUP96 and cytoplasmic filament NUP88^{NTD} (placed in

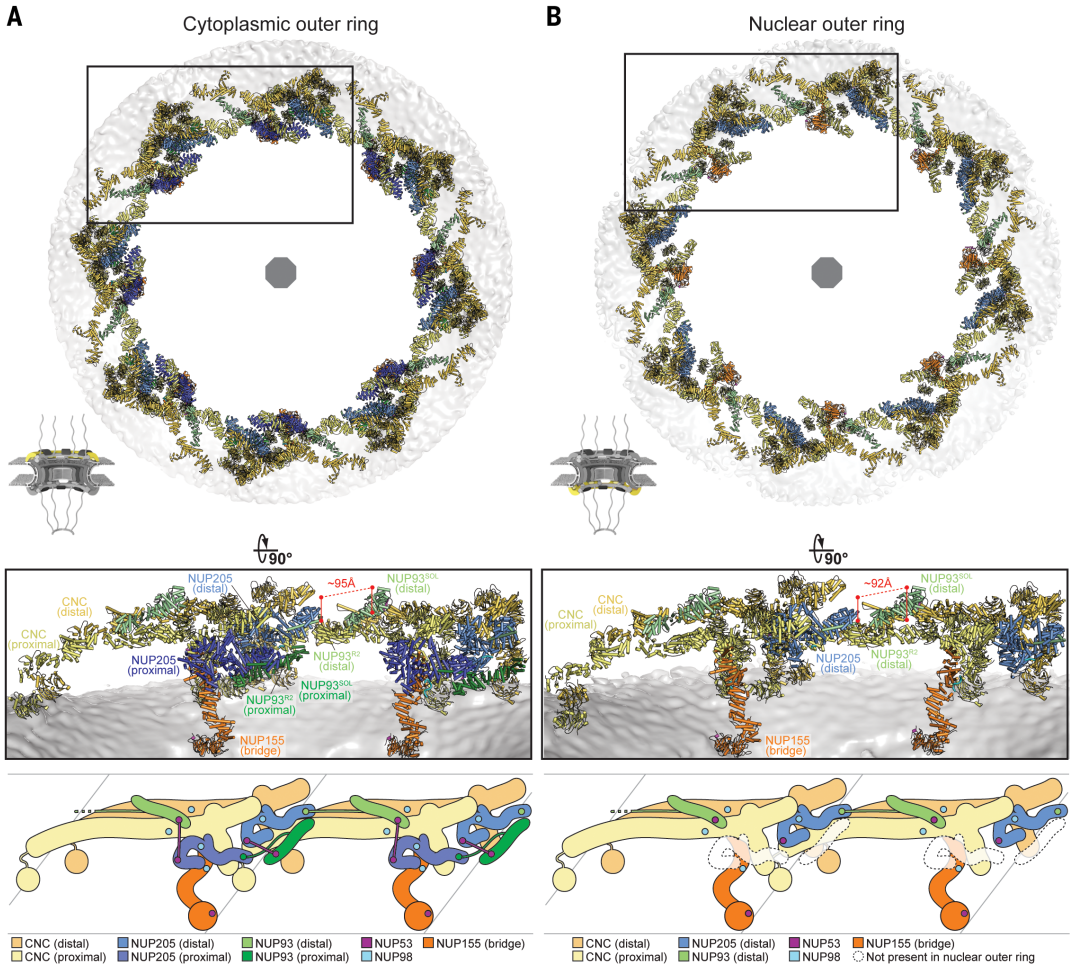


Fig. 7. Architecture of the human NPC symmetric core outer rings. (**A** and **B**) Composite structure generated by quantitatively docking crystal and single-particle cryo-EM structures into an ~12-Å cryo-ET map of the intact human NPC (EMDB ID EMD-14322) (*47*) viewed from the (A) cytoplasmic and (B) nuclear face. Nuclear envelope and docked structures are rendered in isosurface and cartoon representation, respectively. Insets indicate regions encompassing two spokes (top), 90° rotated and magnified (middle), and schematized (bottom). Cross-spoke distances between the distal NUP205-bound NUP93^{R2} and distal NUP93^{SOL} are indicated in red. Linker binding sites on scaffold nup surfaces are indicated by colored circles. Dashed transparent shapes indicate the absence of proximal NUP205 and NUP93 from the nuclear outer ring.

the accompanying manuscript) (*60*) sites are within reach of the ~115-residue linker between the APD and the R1-R2-R3 regions, thus linking the inner ring with the outer ring and cytoplasmic asymmetric portions of the NPC.

The inner ring scaffold architecture is interwoven by linker interactions (Movie 4). The peripheral NUP155-NUP188-NUP93-CNT and equatorial NUP155-NUP205-NUP93-CNT scaffold modules that together form a protomer for a D8-symmetric inner ring can themselves be superposed, if NUP188 and NUP205 are considered as equivalent organizing hubs (Fig. 8G). Nevertheless, the peripheral NUP155 is not linked to the peripheral NUP188•NUP93•CNT complex from the same side of a spoke midplane. Instead, within each spoke, linker-mediated complexes form between a peripheral NUP188•NUP98•NUP93•CNT•NUP53 and an equatorial NUP155 from across the midplane (fig. S82K). On the contrary, the equatorial NUP205•NUP98•NUP93•CNT•NUP53 complex is linked with a peripheral NUP155 from the same and an equatorial NUP155 from the opposite side of a spoke midplane, through NUP98- and NUP53-mediated links, respectively (fig. S82K).

To maximize nup copy parsimony while satisfying all available scaffold binding sites, the human NPC would recruit a total 56 and 80 copies of NUP53 and NUP98, respectively. Though the rest of the symmetric core composite structure agrees with previous estimates of nup stoichiometry, the implied NUP98 and NUP53 copy number exceeds the empirical measurements (*65*). Whereas the discrepancy may be explained by available NUP98 and NUP53 binding sites not being fully occupied, the NUP98 and NUP53 linker nups are known to exchange comparatively rapidly at the NPC (*66*, *67*) and might get depleted as part of the preparation for mass spectrometric analysis.

Docking the composite structure of the NPC into a ~37-Å in situ cryo-ET map of the dilated human NPC revealed that the inner ring spokes move as relatively rigid bodies, accommodating the dilation by increasing the NUP53 linker-bridged gaps between spokes (fig. S83, A and B) (*45*). Spatial restraints in the dilated NPC confirm the intraspoke topology of linkages established by N-terminal NUP93 and NUP98 linkers. Notably, the dilation of the NPC does not induce a substantial increase in the gap between the adjacent spokes of the outer rings, consistent with the cross-linking purported by the linker between distal NUP93 R2 and SOL regions of adjacent spokes (Movie 5 and fig. S83B).

Conclusions

The linker-scaffold is a fundamental architectural principle of the NPC structure. Despite the continuing improvement in resolution attained by cryo-ET reconstructions of intact NPCs over the past decade, the fine molecular detail of the linker-scaffold has remained

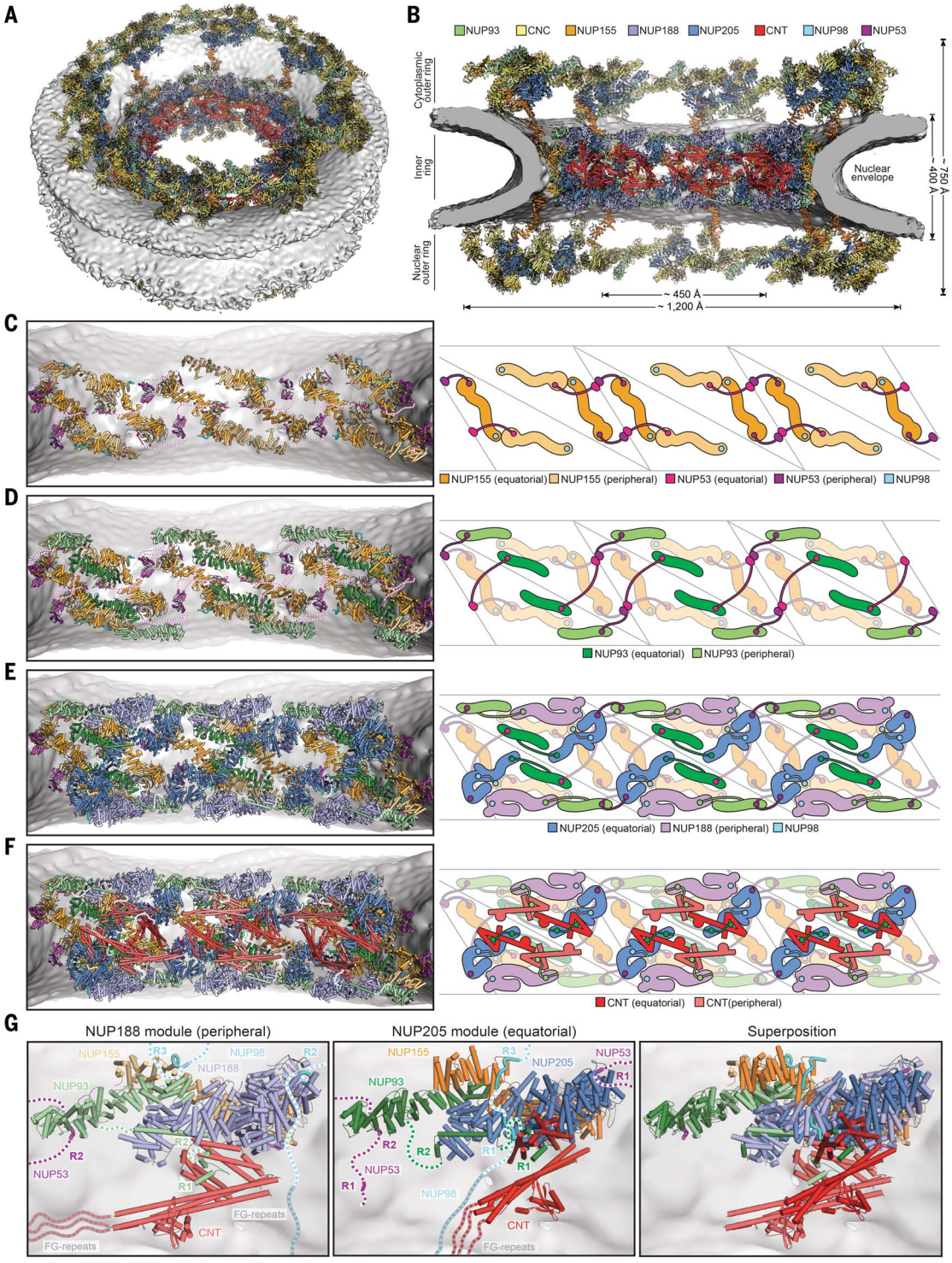
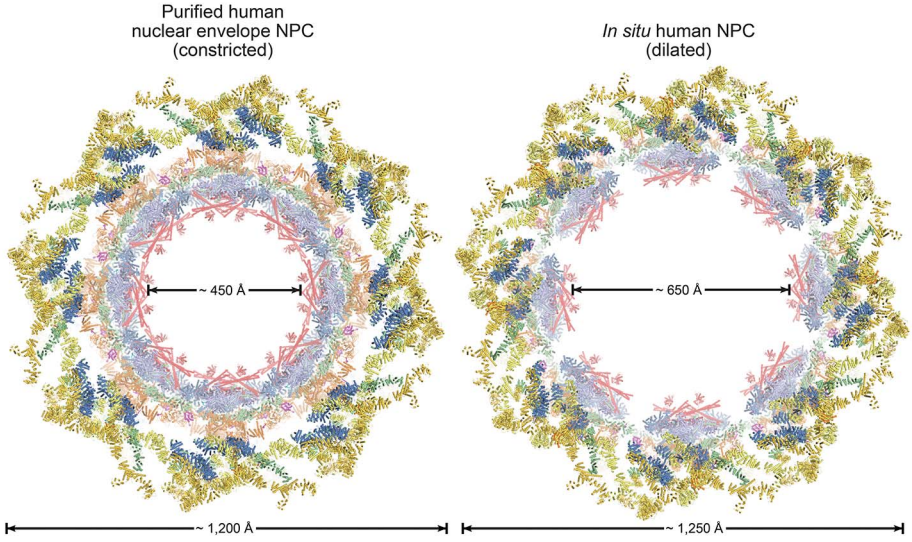


Fig. 8. Linker-scaffold architecture of the human NPC inner ring. (**A** and **B**) Composite structure generated by quantitatively docking crystal and single-particle cryo-EM structures into an ~12-Å cryo-ET map of the intact human NPC (EMDB ID EMD-14322) (*47*) viewed from (A) the cytoplasmic face and (B) the central transport channel cross-section. Nuclear envelope and docked structures are rendered in isosurface and cartoon representation, respectively. (**C** to **F**) Starting from the nuclear envelope, successive layers reveal the architecture of three inner ring spokes of the human NPC. Corresponding schematics illustrate linker paths between binding sites on scaffold surfaces (colored circles). NUP53^{RRM} domains (C) homodimerize between spokes to link cytoplasmic peripheral with nuclear equatorial, and conversely cytoplasmic equatorial with nuclear peripheral copies of NUP155. NUP53^{RRM} domains (D) link cytoplasmic peripheral with nuclear equatorial and, conversely, cytoplasmic equatorial with nuclear peripheral copies of NUP93^{SOL}. NUP205 and NUP188 (E) bind to the equatorial and peripheral NUP93^{R2}, respectively. NUP98 connects NUP205 and equatorial NUP155. NUP53 connects NUP205, and cross-spoke peripheral NUP93^{SOL} (F) CNT is recruited by NUP93^{R1} and positioned by NUP93^{R2} binding to NUP188 and NUP205. (**G**) Close-up views of inner ring modules assembled around NUP188 and NUP205 scaffold hubs and their superposition. Dashed lines indicate unstructured linker nup segments and FG-repeat regions.



Movie 5. Dilation and constriction of the human NPC. Interpolated transition between near-atomic composite structures of the symmetric core in the constricted state of the ~12-Å NPC cryo-ET map (EMDB ID EMD-14322) (*47*) obtained from purified nuclear envelopes and the symmetric core in the dilated state observed in the ~37-Å in situ cryo-ET map (EMDB ID EMD-11967) (*45*) of the human NPC. Enabled by linker-scaffold plasticity, the outward motion of the relatively rigid inner ring spokes enlarges the central transport channel and generates lateral channels between spokes.

out of reach. We used comprehensive residue-level biochemical reconstitution and mapping, crystallographic and single-particle cryo-EM structure determination, in vivo validation, and quantitative docking into improved and diversified cryo-ET NPC reconstructions to delineate the near-atomic structure and evolutionary conservation of the linker-scaffold interactions that underpin the integrity of the NPC.

This study completes the set of structures capturing all biochemically tractable linker-scaffold interactions of the symmetric core. Docked into cryo-ET maps of the human and *S. cerevisiae* NPC, they reveal the topology and restrain distances between linker binding sites on scaffold surfaces, outlining how the multivalent linkers Nup145N/NUP98, Nup53/NUP53 and the N-terminal region of Nic96/NUP93 connect different parts of the NPC. Linkers mediate the formation of inner ring complexes that coalesce into relatively rigid spokes spanning from the nuclear envelope to the central transport channel. They also cross-stitch the inner ring scaffolds with connections across spoke midplanes and flexible links between spokes. In the outer rings, linker-scaffold interactions connect spokes and project ties toward the inner ring.

Our biochemical analysis of linker-scaffold interactions involving Nup145N and Nup53 revealed another architectural principle, invisible to structural methods: Linker-scaffold interactions are driven by structurally defined anchor motifs that present canonical two-component binding dynamics but are potentiated by disperse, structurally elusive interactions between flanking residues and promiscuous binding sites on scaffold surfaces. These Velcro-like binding modes, sometimes referred to as "fuzzy interactions," are found in systems involving intrinsically disordered proteins across biology, a prominent example being the ultrafast exchange of nucleocytoplasmic transport receptors on FG repeats (*12*).

The physical and chemical properties of linkers are advantageous for the assembly of giant complexes like the NPC. The unfolded property of linkers enables long-range interactions and confers flexibility that can accommodate large movements or shock-absorb nuclear envelope deformations. The disperse nature of linker-scaffold interactions is conducive to the reuse of linker interactions in different chemical and steric environments of the NPC. The ensemble of binding modes provides robustness in the face of conformational changes of the NPC that might otherwise be incompatible with a singular binding mode. The bulk of FG repeats present in the central transport channel, which form promiscuous transient interactions with the inner ring scaffold and other parts of the NPC, is likely to have a similar effect. Specifically, the previous findings that FG repeats not only form the NPC's diffusion barrier but also interact with scaffolds supports this notion (*49*, *52*, *58*, *68*, *69*). Considering the avidity that results from multiple scaffold valences per linker, and the allovalency mediated by flanking residues (*70*, *71*), it is unsurprising that our in vivo perturbation of interactions required extensive mutations to exacerbate deleterious phenotypes. For this reason and because of the lack of constraints imposed by a protein fold, new linker sequences are readily evolvable. Notably, the linker-scaffold network topology and modular binding site distribution on linkers is conserved from fungi to humans, despite considerable divergence in linker sequences, most extremely exemplified by the complete divergence of the Nup53/NUP53 motif that binds to a conserved site on Nic96/NUP93.

The binding of linkers is amenable to exchange and regulation. The linearity of linkers imposes few obstacles to the deposition of posttranslational modifications by the same machinery along the entire sequence to rapidly ablate the multiple binding valences. Indeed, patterns of Cdk1 and Nek-driven phosphorylation that lead to the choreographed depletion of both NUP98 and NUP53 from the NPC during mitotic nuclear envelope breakdown include the R1, R2, and R3 regions of both linkers (*72*, *73*). Structural defects in the NPC resulting in aberrant nucleocytoplasmic transport may affect gene expression, mRNA maturation, and mRNA export, leading to downstream tumorigenic processes. Therefore, the depletion of NUP98 from the NPC as a result of gene fusion mutations associated with various hematopoietic malignancies should be considered in the study of the carcinogenic mechanisms triggered by these mutations (*74*).

The unexpected discovery of the presence and distinctive role of NUP93 in cross-linking the outer ring spokes of the human NPC, along with its organizing role in the inner ring as both scaffold and linker, exemplifies the reuse of linker-scaffold functional units at completely different locations of the NPC. Its ubiquity rationalizes the observation that rapid degron-induced depletion of NUP93 leads to the concomitant loss of both inner and outer rings from the nuclear pore, legitimizing NUP93 as a "lynchpin" of the NPC (*75*). These findings further inform the mechanistic basis for pathologies like steroid-resistant nephrotic syndrome (SRNS), which is associated with mutations in NUP93 and NUP205, the most poignant of which is NUP205 Phe^{1995}→Ser (F1995S), located in the NUP93^{R2} binding site of NUP205 and shown to abolish the NUP205-NUP93 interaction (*76*). Nonsense mutations that omit the NUP93^{R2} binding site in NUP188 are also associated with neurologic, ocular, and cardiac abnormalities (*77*).

The docking of our NPC composite structure into an ~37-Å in situ cryo-ET map of the dilated human NPC demonstrates the magnitude of the movements that must be withstood by the linkers that connect the inner ring spokes (*45*). Importantly, the dilated inner ring reveals lateral channels between its spokes that can

accommodate the passage of small cytosolic domains of inner nuclear membrane integral membrane proteins (INM-IMPs), suggesting the mechanism by which inner ring constriction upon energy depletion might interfere with the path of diffusion of these proteins between outer and inner nuclear envelope membranes (*78*–*80*). It remains to be established whether the extent of inner ring dilation observed by current cryo-ET reconstructions of the human and *S. cerevisiae* NPCs capture the maximally achievable lateral channel dimensions. However, the previous observations that karyopherin-mediated active INM-IMP transport requires unstructured tethers spanning the distance between the nuclear envelope and the central transport channel is consistent with the observed lateral channel dimensions (fig. S83C) (*81*–*85*). Taking advantage of the composite NPC structures, future studies are expected to elucidate the nucleocytoplasmic translocation pathways and the impact of inner ring dilation for individual INM-IMPs. Similarly, the translocation of perinuclear domains on the opposite side of the nuclear envelope is expected to be limited by the luminal ring that encircles the NPC midplane formed by Pom152/POM210 Ig-like domains (*37*, *47*, *86*).

Finally, the capacity of the NPC to exist in dilated and constricted states dependent on tension imparted by the surrounding nuclear envelope portrays the NPC as the cell's largest mechanosensitive channel, with implications in cellular energy–state sensing and transport of transcriptional regulators in response to mechanical stress on the cell (*46*, *87*). The constriction and dilation of the inner ring may not only affect the distribution and local concentration of FG repeats in the central transport channel, a determinant of karyopherin-mediated transport efficiency (*88*) but also sterically modulate the flux of INM-IMPs and large cargos such as preribosome or messenger ribonucleoprotein particles (mRNPs). Future studies will have to establish the causal links between the transmission of tension from the nuclear envelope to the NPC, the dilation of its inner ring, and mechanisms by which dilation and constriction may modulate nucleocytoplasmic transport.

Our results illuminate the elusive linker-scaffold molecular interactions that maintain the integrity of the NPC, providing a comprehensive characterization of a final major aspect of the NPC symmetric core. Building on this roadmap, future studies can address NPC assembly and disassembly mechanisms, the emergence of NPC polarity and attachment of the asymmetric cytoplasmic filaments and nuclear basket to the symmetric core, mechanisms of NPC-associated diseases, and mechanisms of NPC dilation and constriction along with their implications in nucleocytoplasmic transport.

Methods summary

Comprehensive materials and methods are presented in the supplementary materials. Briefly, the source of materials and reagents is listed in table S1. Summaries of bacterial expression constructs and conditions (table S2), protein purification procedures (table S3), analytical SEC-MALS protein interaction analyses (table S4), and ITC binding affinity measurements (table S5) are provided. Experimental details of x-ray crystallography and single-particle cryo-EM structure determination procedures are described, including summaries of crystallization and cryo-protection conditions (table S6), as well as data collection, processing, and refinement statistics (tables S7 to S12). *S. cerevisiae* constructs (table S13) and strains (table S14), as well as the experimental details of the viability and growth assay, subcellular nup localization analysis, *60S* preribosome export assay, and mRNA export fluorescence in situ hybridization (FISH) assay, which establish the physiological relevance of the biochemical and structural findings, are provided. Details of the incremental quantitative docking procedures for nup and nup complex crystal and single-particle cryo-EM structures into ~12- and ~23-Å cryo-ET maps of the intact human NPC (constricted state) (*38*, *47*), as well as into an ~37-Å in situ cryo-ET map of the human NPC (*45*) and an ~25-Å in situ map of the *S. cerevisiae* NPC (*36*) (dilated states), are provided. Inventories of nup and nup complex experimental structures used to generate the near-atomic composite structures of the *S. cerevisiae* (table S15) and human (table S16) NPCs are provided.

REFERENCES AND NOTES

1. A. Hoelz, E. W. Debler, G. Blobel, The structure of the nuclear pore complex. *Annu. Rev. Biochem.* **80**, 613–643 (2011). doi: 10.1146/annurev-biochem-060109-151030; pmid: 21495847
2. D. H. Lin, A. Hoelz, The structure of the nuclear pore complex (an update). *Annu. Rev. Biochem.* **88**, 725–783 (2019). doi: 10.1146/annurev-biochem-062917-011901; pmid: 30883195
3. B. Hampoelz, A. Andres-Pons, P. Kastritis, M. Beck, Structure and assembly of the nuclear pore complex. *Annu. Rev. Biophys.* **48**, 515–536 (2019). doi: 10.1146/annurev-biophys-052118-115308; pmid: 30943044
4. K. E. Knockenhauer, T. U. Schwartz, The nuclear pore complex as a flexible and dynamic gate. *Cell* **164**, 1162–1171 (2016). doi: 10.1016/j.cell.2016.01.034; pmid: 26967283
5. A. Köhler, E. Hurt, Gene regulation by nucleoporins and links to cancer. *Mol. Cell* **38**, 6–15 (2010). doi: 10.1016/j.molcel.2010.01.040; pmid: 20385085
6. M. L. Yarbrough, M. A. Mata, R. Sakthivel, B. M. Fontoura, Viral subversion of nucleocytoplasmic trafficking. *Traffic* **15**, 127–140 (2014). doi: 10.1111/tra.12137; pmid: 24289861
7. P. L. Paine, L. C. Moore, S. B. Horowitz, Nuclear envelope permeability. *Nature* **254**, 109–114 (1975). doi: 10.1038/254109a0; pmid: 1117994
8. W. M. Bonner, Protein migration into nuclei. I. Frog oocyte nuclei in vivo accumulate microinjected histones, allow entry to small proteins, and exclude large proteins. *J. Cell Biol.* **64**, 421–430 (1975). doi: 10.1083/jcb.64.2.421; pmid: 46868
9. B. L. Timney *et al*., Simple rules for passive diffusion through the nuclear pore complex. *J. Cell Biol.* **215**, 57–76 (2016). doi: 10.1083/jcb.201601004; pmid: 27697925
10. Y. M. Chook, K. E. Süel, Nuclear import by karyopherin-βs: Recognition and inhibition. *Biochim. Biophys. Acta* **1813**, 1593–1606 (2011). doi: 10.1016/j.bbamcr.2010.10.014; pmid: 21029754
11. A. Cook, F. Bono, M. Jinek, E. Conti, Structural biology of nucleocytoplasmic transport. *Annu. Rev. Biochem.* **76**, 647–671 (2007). doi: 10.1146/annurev.biochem.76.052705.161529; pmid: 17506639
12. S. Milles *et al*., Plasticity of an ultrafast interaction between nucleoporins and nuclear transport receptors. *Cell* **163**, 734–745 (2015). doi: 10.1016/j.cell.2015.09.047; pmid: 26456112
13. R. A. Pumroy, G. Cingolani, Diversification of importin-α isoforms in cellular trafficking and disease states. *Biochem. J.* **466**, 13–28 (2015). doi: 10.1042/BJ20141186; pmid: 25656054
14. A. E. Hodel *et al*., The three-dimensional structure of the autoproteolytic, nuclear pore-targeting domain of the human nucleoporin Nup98. *Mol. Cell* **10**, 347–358 (2002). doi: 10.1016/S1097-2765(02)00589-0; pmid: 12191480
15. I. C. Berke, T. Boehmer, G. Blobel, T. U. Schwartz, Structural and functional analysis of Nup133 domains reveals modular building blocks of the nuclear pore complex. *J. Cell Biol.* **167**, 591–597 (2004). doi: 10.1083/jcb.200408109; pmid: 15557116
16. K. C. Hsia, P. Stavropoulos, G. Blobel, A. Hoelz, Architecture of a coat for the nuclear pore membrane. *Cell* **131**, 1313–1326 (2007). doi: 10.1016/j.cell.2007.11.038; pmid: 18160040
17. T. Boehmer, S. Jeudy, I. C. Berke, T. U. Schwartz, Structural and functional studies of Nup107/Nup133 interaction and its implications for the architecture of the nuclear pore complex. *Mol. Cell* **30**, 721–731 (2008). doi: 10.1016/j.molcel.2008.04.022; pmid: 18570875
18. S. G. Brohawn, N. C. Leksa, E. D. Spear, K. R. Rajashankar, T. U. Schwartz, Structural evidence for common ancestry of the nuclear pore complex and vesicle coats. *Science* **322**, 1369–1373 (2008). doi: 10.1126/science.1165886; pmid: 18974315
19. E. W. Debler *et al*., A fence-like coat for the nuclear pore membrane. *Mol. Cell* **32**, 815–826 (2008). doi: 10.1016/j.molcel.2008.12.001; pmid: 19111661
20. S. G. Brohawn, T. U. Schwartz, Molecular architecture of the Nup84-Nup145C-Sec13 edge element in the nuclear pore complex lattice. *Nat. Struct. Mol. Biol.* **16**, 1173–1177 (2009). doi: 10.1038/nsmb.1713; pmid: 19855394
21. N. C. Leksa, S. G. Brohawn, T. U. Schwartz, The structure of the scaffold nucleoporin Nup120 reveals a new and unexpected domain architecture. *Structure* **17**, 1082–1091 (2009). doi: 10.1016/j.str.2009.06.003; pmid: 19576787
22. V. Nagy *et al*., Structure of a trimeric nucleoporin complex reveals alternate oligomerization states. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 17693–17698 (2009). doi: 10.1073/pnas.0909373106; pmid: 19805193
23. H. S. Seo *et al*., Structural and functional analysis of Nup120 suggests ring formation of the Nup84 complex. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14281–14286 (2009). doi: 10.1073/pnas.0907453106; pmid: 19706512
24. J. R. R. Whittle, T. U. Schwartz, Architectural nucleoporins Nup157/170 and Nup133 are structurally related and descend from a second ancestral element. *J. Biol. Chem.* **284**, 28442–28452 (2009). doi: 10.1074/jbc.M109.023580; pmid: 19674973
25. S. Bilokapic, T. U. Schwartz, Molecular basis for Nup37 and ELY5/ELYS recruitment to the nuclear pore complex. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15241–15246 (2012). doi: 10.1073/pnas.1205151109; pmid: 22955883
26. S. Bilokapic, T. U. Schwartz, Structural and functional studies of the 252 kDa nucleoporin ELYS reveal distinct roles for its three tethered domains. *Structure* **21**, 572–580 (2013). doi: 10.1016/j.str.2013.02.006; pmid: 23499022
27. K. Kelley, K. E. Knockenhauer, G. Kabachinski, T. U. Schwartz, Atomic structure of the Y complex of the nuclear pore. *Nat. Struct. Mol. Biol.* **22**, 425–431 (2015). doi: 10.1038/nsmb.2998; pmid: 25822992
28. T. Stuwe *et al*., Architecture of the fungal nuclear pore inner ring complex. *Science* **350**, 56–64 (2015). doi: 10.1126/science.aac9176; pmid: 26316600
29. T. Stuwe *et al*., Architecture of the nuclear pore complex coat. *Science* **347**, 1148–1152 (2015). doi: 10.1126/science.aaa4136; pmid: 25745173
30. C. Xu *et al*., Crystal structure of human nuclear pore complex component NUP43. *FEBS Lett.* **589**, 3247–3253 (2015). doi: 10.1016/j.febslet.2015.09.008; pmid: 26391640
31. A. Hoelz, J. S. Glavy, M. Beck, Toward the atomic structure of the nuclear pore complex: When top down meets bottom up. *Nat. Struct. Mol. Biol.* **23**, 624–630 (2016). doi: 10.1038/nsmb.3244; pmid: 27273515
32. K. H. Bui *et al*., Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* **155**, 1233–1243 (2013). doi: 10.1016/j.cell.2013.10.055; pmid: 24315095

33. J. Fischer, R. Teimer, S. Amlacher, R. Kunze, E. Hurt, Linker Nups connect the nuclear pore complex inner ring with the outer ring and transport channel. *Nat. Struct. Mol. Biol.* **22**, 774–781 (2015). doi: 10.1038/nsmb.3084; pmid: 26344569
34. J. Kosinski *et al*., Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352**, 363–365 (2016). doi: 10.1126/science.aaf0643; pmid: 27081072
35. D. H. Lin *et al*., Architecture of the symmetric core of the nuclear pore. *Science* **352**, aaf1015 (2016). doi: 10.1126/science.aaf1015; pmid: 27081075
36. M. Allegretti *et al*., In-cell architecture of the nuclear pore and snapshots of its turnover. *Nature* **586**, 796–800 (2020). doi: 10.1038/s41586-020-2670-5; pmid: 32879490
37. S. J. Kim *et al*., Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475–482 (2018). doi: 10.1038/nature26003; pmid: 29539637
38. A. von Appen *et al*., In situ structural analysis of the human nuclear pore complex. *Nature* **526**, 140–143 (2015). doi: 10.1038/nature15381; pmid: 26416747
39. S. A. Nordeen, D. L. Turman, T. U. Schwartz, Yeast Nup84-Nup133 complex structure details flexibility and reveals conservation of the membrane anchoring ALPS motif. *Nat. Commun.* **11**, 6060 (2020). doi: 10.1038/s41467-020-19885-5; pmid: 33247142
40. S. Amlacher *et al*., Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. *Cell* **146**, 277–289 (2011). doi: 10.1016/j.cell.2011.06.039; pmid: 21784248
41. T. Stuwe, D. H. Lin, L. N. Collins, E. Hurt, A. Hoelz, Evidence for an evolutionary relationship between the large adaptor nucleoporin Nup192 and karyopherins. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2530–2535 (2014). doi: 10.1073/pnas.1311081111; pmid: 24505056
42. H. Chug, S. Trakhanov, B. B. Hülsmann, T. Pleiner, D. Görlich, Crystal structure of the metazoan Nup62•Nup58•Nup54 nucleoporin complex. *Science* **350**, 106–110 (2015). doi: 10.1126/science.aac7420; pmid: 26292704
43. S. Mosalaganti *et al*., In situ architecture of the algal nuclear pore complex. *Nat. Commun.* **9**, 2361 (2018). doi: 10.1038/s41467-018-04739-y; pmid: 29915221
44. A. P. Schuller *et al*., The cellular environment shapes the nuclear pore complex architecture. *Nature* **598**, 667–671 (2021). doi: 10.1038/s41586-021-03985-3; pmid: 34646014
45. V. Zila *et al*., Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell* **184**, 1032–1046.e18 (2021). doi: 10.1016/j.cell.2021.01.025; pmid: 33571428
46. C. E. Zimmerli *et al*., Nuclear pores dilate and constrict in cellulo. *Science* **374**, eabd9776 (2021). doi: 10.1126/science.abd9776; pmid: 34762489
47. S. Mosalaganti *et al*., AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science* **376**, eabm9506 (2022). doi: 10.1126/science.abm9506
48. P. Sampathkumar *et al*., Structure, dynamics, evolution, and function of a major scaffold component in the nuclear pore complex. *Structure* **21**, 560–571 (2013). doi: 10.1016/j.str.2013.02.005; pmid: 23499021
49. K. R. Andersen *et al*., Scaffold nucleoporins Nup188 and Nup192 share structural and functional properties with nuclear transport receptors. *eLife* **2**, e00745 (2013). doi: 10.7554/eLife.00745; pmid: 23795296
50. S. M. Bailer *et al*., Nup116p and nup100p are interchangeable through a conserved motif which constitutes a docking site for the mRNA transport factor gle2p. *EMBO J.* **17**, 1107–1119 (1998). doi: 10.1093/emboj/17.4.1107; pmid: 9463388
51. K. J. Ryan, S. R. Wente, Isolation and characterization of new Saccharomyces cerevisiae mutants perturbed in nuclear pore complex assembly. *BMC Genet.* **3**, 17 (2002). doi: 10.1186/1471-2156-3-17; pmid: 12215173
52. S. S. Patel, B. J. Belmont, J. M. Sante, M. F. Rexach, Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. *Cell* **129**, 83–96 (2007). doi: 10.1016/j.cell.2007.01.044; pmid: 17418788
53. K. Yoshida, H. S. Seo, E. W. Debler, G. Blobel, A. Hoelz, Structural and functional analysis of an essential nucleoporin heterotrimer on the cytoplasmic face of the nuclear pore complex. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16571–16576 (2011). doi: 10.1073/pnas.1112846108; pmid: 21930948
54. A. de Bruyn Kops, C. Guthrie, Identification of the novel *Nup188-brr7* allele in a screen for cold-sensitive mRNA export mutants in *Saccharomyces cerevisiae*. *G3* **8**, 2991–3003 (2018). doi: 10.1534/g3.118.200447; pmid: 30021831
55. M. Miao, K. J. Ryan, S. R. Wente, The integral membrane protein Pom34p functionally links nucleoporin subcomplexes. *Genetics* **172**, 1441–1457 (2006). doi: 10.1534/genetics.105.052068; pmid: 16361228
56. U. Nehrbass, M. P. Rout, S. Maguire, G. Blobel, R. W. Wozniak, The yeast nucleoporin Nup188p interacts genetically and physically with the core structures of the nuclear pore complex. *J. Cell Biol.* **133**, 1153–1162 (1996). doi: 10.1083/jcb.133.6.1153; pmid: 8682855
57. M. Marelli, J. D. Aitchison, R. W. Wozniak, Specific binding of the karyopherin Kap121p to a subunit of the nuclear pore complex containing Nup53p, Nup59p, and Nup170p. *J. Cell Biol.* **143**, 1813–1830 (1998). doi: 10.1083/jcb.143.7.1813; pmid: 9864357
58. N. Schrader *et al*., Structural basis of the nic96 subcomplex organization in the nuclear pore channel. *Mol. Cell* **29**, 46–55 (2008). doi: 10.1016/j.molcel.2007.10.022; pmid: 18206968
59. S. Jeudy, T. U. Schwartz, Crystal structure of nucleoporin Nic96 reveals a novel, intricate helical domain architecture. *J. Biol. Chem.* **282**, 34904–34912 (2007). doi: 10.1074/jbc.M705479200; pmid: 17897938
60. C. J. Bley *et al*., Architecture of the cytoplasmic face of the nuclear pore. *Science* **376**, eabm9129 (2022). doi: 10.1126/science.abm9129
61. D. H. Lin *et al*., Structural and functional analysis of mRNA export regulation by the nuclear pore complex. *Nat. Commun.* **9**, 2319 (2018). doi: 10.1038/s41467-018-04459-3; pmid: 29899397
62. G. Drin *et al*., A general amphipathic alpha-helical motif for sensing membrane curvature. *Nat. Struct. Mol. Biol.* **14**, 138–146 (2007). doi: 10.1038/nsmb1194; pmid: 17220896
63. N. Eisenhardt, J. Redolfi, W. Antonin, Interaction of Nup53 with Ndc1 and Nup155 is required for nuclear pore complex assembly. *J. Cell Sci.* **127**, 908–921 (2014). pmid: 24363447
64. E. Onischenko, L. H. Stanton, A. S. Madrid, T. Kieselbach, K. Weis, Role of the Ndc1 interaction network in yeast nuclear pore complex assembly and maintenance. *J. Cell Biol.* **185**, 475–491 (2009). doi: 10.1083/jcb.200810030 pmid: 19414609
65. A. Ori *et al*., Cell type-specific nuclear pores: A case in point for context-dependent stoichiometry of molecular machines. *Mol. Syst. Biol.* **9**, 648 (2013). doi: 10.1038/msb.2013.4; pmid: 23511206
66. E. R. Griffis, N. Altan, J. Lippincott-Schwartz, M. A. Powers, Nup98 is a mobile nucleoporin with transcription-dependent dynamics. *Mol. Biol. Cell* **13**, 1282–1297 (2002). doi: 10.1091/mbc.01-11-0538; pmid: 11950939
67. G. Rabut, V. Doye, J. Ellenberg, Mapping the dynamic organization of the nuclear pore complex inside single living cells. *Nat. Cell Biol.* **6**, 1114–1121 (2004). doi: 10.1038/ncb1184; pmid: 15502822
68. N. P. Allen, L. Huang, A. Burlingame, M. Rexach, Proteomic analysis of nucleoporin interacting proteins. *J. Biol. Chem.* **276**, 29268–29274 (2001). doi: 10.1074/jbc.M102629200; pmid: 11387327
69. E. Onischenko *et al*., Natively unfolded FG repeats stabilize the structure of the nuclear pore complex. *Cell* **171**, 904–917.e19 (2017). doi: 10.1016/j.cell.2017.09.033; pmid: 29033133
70. A. Levchenko, Allovalency: A case of molecular entanglement. *Curr. Biol.* **13**, R876–R878 (2003). doi: 10.1016/j.cub.2003.10.049; pmid: 14614843
71. J. G. Olsen, K. Teilum, B. B. Kragelund, Behaviour of intrinsically disordered proteins in protein-protein complexes with an emphasis on fuzziness. *Cell. Mol. Life Sci.* **74**, 3175–3183 (2017). doi: 10.1007/s00018-017-2560-7; pmid: 28597296
72. E. Laurell *et al*., Phosphorylation of Nup98 by multiple kinases is crucial for NPC disassembly during mitotic entry. *Cell* **144**, 539–550 (2011). doi: 10.1016/j.cell.2011.01.012; pmid: 21335236
73. M. I. Linder *et al*., Mitotic disassembly of nuclear pore complexes involves CDK1- and PLK1-mediated phosphorylation of key interconnecting nucleoporins. *Dev. Cell* **43**, 141–156.e7 (2017). doi: 10.1016/j.devcel.2017.08.020; pmid: 29065306
74. S. M. Gough, C. I. Slape, P. D. Aplan, NUP98 gene fusions and hematopoietic malignancies: Common themes and new biologic insights. *Blood* **118**, 6247–6257 (2011). doi: 10.1182/blood-2011-07-328880; pmid: 21948299
75. S. G. Regmi *et al*., The nuclear pore complex consists of two independent scaffolds. bioRxiv 2020.11.13.381947 [Preprint] (2020); https://doi.org/10.1101/2020.11.13.381947.
76. D. A. Braun *et al*., Mutations in nuclear pore genes *NUP93*, *NUP205* and *XPO5* cause steroid-resistant nephrotic syndrome. *Nat. Genet.* **48**, 457–465 (2016). doi: 10.1038/ng.3512; pmid: 26878725
77. A. M. Muir *et al*., Bi-allelic loss-of-function variants in *NUP188* cause a recognizable syndrome characterized by neurologic, ocular, and cardiac abnormalities. *Am. J. Hum. Genet.* **106**, 623–631 (2020). doi: 10.1016/j.ajhg.2020.03.009; pmid: 32275884
78. R. Ungricht, M. Klann, P. Horvath, U. Kutay, Diffusion and retention are major determinants of protein targeting to the inner nuclear membrane. *J. Cell Biol.* **209**, 687–704 (2015). doi: 10.1083/jcb.201409127; pmid: 26056139
79. T. Ohba, E. C. Schirmer, T. Nishimoto, L. Gerace, Energy- and temperature-dependent transport of integral proteins to the inner nuclear membrane via the nuclear pore. *J. Cell Biol.* **167**, 1051–1062 (2004). doi: 10.1083/jcb.200409149; pmid: 15611332
80. N. Zuleger *et al*., System analysis shows distinct mechanisms and common principles of nuclear envelope protein dynamics. *J. Cell Biol.* **193**, 109–123 (2011). doi: 10.1083/jcb.201009068; pmid: 21444689
81. M. C. King, C. P. Lusk, G. Blobel, Karyopherin-mediated import of integral inner nuclear membrane proteins. *Nature* **442**, 1003–1007 (2006). doi: 10.1038/nature05075; pmid: 16929305
82. A. C. Meinema *et al*., Long unfolded linkers facilitate membrane protein import through the nuclear pore complex. *Science* **333**, 90–93 (2011). doi: 10.1126/science.1205741; pmid: 21659568
83. R. K. Lokareddy *et al*., Distinctive properties of the nuclear localization signals of inner nuclear membrane proteins Heh1 and Heh2. *Structure* **23**, 1305–1316 (2015). doi: 10.1016/j.str.2015.04.017; pmid: 26051712
84. P. Popken, A. Ghavami, P. R. Onck, B. Poolman, L. M. Veenhoff, Size-dependent leak of soluble and membrane proteins through the yeast nuclear pore complex. *Mol. Biol. Cell* **26**, 1386–1394 (2015). doi: 10.1091/mbc.E14-07-1175; pmid: 25631821
85. A. Kralt *et al*., Conservation of inner nuclear membrane targeting sequences in mammalian Pom121 and yeast Heh2 membrane proteins. *Mol. Biol. Cell* **26**, 3301–3312 (2015). doi: 10.1091/mbc.e15-03-0184; pmid: 26179916
86. Y. Zhang *et al*., Molecular architecture of the luminal ring of the *Xenopus laevis* nuclear pore complex. *Cell Res.* **30**, 532–540 (2020). doi: 10.1038/s41422-020-0320-y; pmid: 32367042
87. A. Elosegui-Artola *et al*., Force triggers YAP nuclear entry by regulating transport across nuclear pores. *Cell* **171**, 1397–1410.e14 (2017). doi: 10.1016/j.cell.2017.10.008; pmid: 29107331
88. S. Frey, D. Görlich, A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes. *Cell* **130**, 512–523 (2007). doi: 10.1016/j.cell.2007.06.024; pmid: 17693259
89. S. Petrovic *et al*., Architecture of the linker-scaffold in the nuclear pore, Version 1.0. CaltechDATA (2021); https://doi.org/10.22002/D1.2208.

ACKNOWLEDGMENTS

We thank A. Patke for critical reading and editing of the manuscript and insightful discussions; M. Beck for sharing an ~12-Å cryo-ET reconstruction of the intact human HeLa cell NPC before publication [Electron Microscopy Data Bank (EMDB) ID EMD-14322]; E. Hurt, S. Wente, B. Fountura, D. Baltimore, and the Kazusa DNA Research Institute for providing material; and F. Liang, A. Lyons, A. Tang, and J. Thai for experimental support. We are grateful to D. Borek, J. Kollman, G. Lander, D. Lin, S. Saladi, and members of the Hoelz lab for insightful discussion and expertise. We acknowledge J. Kaiser, the scientific staff of the Stanford Synchrotron Radiation Lightsource (SSRL) beamline 12-2, and the National Institute of General Medical Sciences and National Cancer Institute Structural Biology Facility (GM/CA) at the Advanced Photon Source (APS) for their support with x-ray diffraction measurements; S. Chen and A. Malyutin of the Beckman Institute Resource Center for Transmission Electron Microscopy at the California Institute of Technology (Caltech) for support with cryo-EM microscopy imaging; J. Myers and the scientific staff of the Pacific Northwest CryoEM Center (PNCC) at the Oregon Health and Science University (OHSU) and the Environmental Molecular Sciences Laboratory (EMSL) for their support with cryo-EM imaging; and the Cold Spring Harbor Laboratory (CSHL) cryo-EM course, the CSHL X-ray Methods in Structural Biology course, and the Michigan Life Science Institute cryo-EM workshop and their instructors M. Cianfrocco, W. Furey, G. Gilliland, J. Kollman, G. Lander, M. Ohi, J. Pflugrath, A. McPherson, and M. Vos, along with all the course staff and lecturers for valuable expert training. **Funding:** The Molecular Observatory at Caltech is supported by D. and J. Voet, the Gordon and Betty Moore Foundation, and the Beckman Institute. The Center for Molecular Medicine at Caltech is supported by the Gordon and Betty Moore Foundation. The operations at the SSRL and APS are supported by the US Department of Energy (DOE) and the National Institutes of Health

(NIH). GM/CA has been funded in whole or in part with federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006). A portion of this research was supported by NIH grant U24GM129547 and performed at the PNCC at OHSU and accessed through EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research. A.H. was supported by a Camille-Dreyfus Teacher Scholar Award (TC-15-082) and NIH grants R01-GM117360 and R01-GM111461, is an investigator of the Heritage Medical Research Institute (HMRI-15-09-01), and is a faculty scholar of the Howard Hughes Medical Institute (55108534). S.P. was supported by a predoctoral fellowship from the Boehringer Ingelheim Fonds and by an Amgen Graduate Fellowship through the Caltech-Amgen Research Collaboration. **Author contributions:** A.H. conceived and coordinated the study. S.P., D.S., T.P., C.J.B., K.T., and A.H. designed the research. S.P., D.S., T.P., C.J.B., K.T., B.B., G.P.T., and L.S. performed the research. S.P., D.S., T.P., C.J.B., K.T., B.B., S.N., G.W.M., T.A.S., X.L., G.P.T., L.S., and A.H. analyzed data. S.P., D.S., T.P., C.J.B., and A.H. integrated and conceptualized the results. D.S. and T.P. contributed momentously and equally to this work. S.P., C.J.B., S.N., G.W.M., and A.H. wrote and revised the manuscript, with contributions from all authors. **Competing interests:** The authors declare no conflicts of interest. **Data and materials availability:** Materials generated in this study are available on request from the corresponding author. The coordinates and structure factors of crystal structures have been deposited in the Protein Data Bank (PDB) with accession numbers 7MVT (Nup192$^{\Delta Head}$•Nic96$^{187\text{-}301}$), 7MVW (Nup188^{NTD}), 7MVX (Nup188•Nic96^{R2}), 7MW0 (NUP93^{SOL}), and 7MW0 (NUP93^{SOL}•NUP53^{R2}). The coordinates of single particle cryo-EM structures have been deposited in the PDB with accession MVU (Nup192•Nic96^{R2}), 7MVV (Nup192•Nic96^{R2}•Nup145N^{R1}•Nup53^{R1}), 7MVY (Nup188•Nic96^{R2}), and 7MVZ (Nup188•Nic96^{R2}•Nup145N^{R2}). The maps of single particle cryo-EM structures have been deposited in the EMDB with accession numbers EMD-24056 (Nup192•Nic96^{R2}), EMD-24057 (Nup192•Nic96^{R2}•Nup145N^{R1}•Nup53^{R1}), EMD-24058 (Nup188•Nic96^{R2}), and EMD-24059 (Nup188•Nic96^{R2}•Nup145N^{R2}). PyMol and Chimera sessions containing the composite structures of the constricted human, dilated human, and dilated *S. cerevisiae* NPC symmetric core can be obtained from our webpage (http://ahweb.caltech.edu), and coordinates are deposited in the PDB with the respective accession numbers 7TBJ, 7TBK, and 7TBI. Quantitative docking data, workflow code, PyMol, and Chimera sessions were deposited on CaltechDATA (*89*). **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.abm9798
Materials and Methods
Supplementary Text
Figs. S1 to S83
Tables S1 to S16
References (*90*–*148*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 26 October 2021; accepted 15 April 2022
10.1126/science.abm9798

RESEARCH ARTICLE

QUANTUM COMPUTING

Quantum advantage in learning from experiments

Hsin-Yuan Huang[1,2]*, Michael Broughton[3], Jordan Cotler[4,5], Sitan Chen[6,7], Jerry Li[8], Masoud Mohseni[3], Hartmut Neven[3], Ryan Babbush[3], Richard Kueng[9], John Preskill[1,2,10], Jarrod R. McClean[3]*

Quantum technology promises to revolutionize how we learn about the physical world. An experiment that processes quantum data with a quantum computer could have substantial advantages over conventional experiments in which quantum states are measured and outcomes are processed with a classical computer. We proved that quantum machines could learn from exponentially fewer experiments than the number required by conventional experiments. This exponential advantage is shown for predicting properties of physical systems, performing quantum principal component analysis, and learning about physical dynamics. Furthermore, the quantum resources needed for achieving an exponential advantage are quite modest in some cases. Conducting experiments with 40 superconducting qubits and 1300 quantum gates, we demonstrated that a substantial quantum advantage is possible with today's quantum processors.

Humans learn about nature through experiments, but until now our ability to acquire knowledge has been hindered by viewing the quantum world through a classical lens. The rapid advancement of quantum technology portends an opportunity to observe the world in a fundamentally different and more powerful way. Instead of measuring physical systems and then processing the classical measurement outcomes to infer properties of those physical systems, quantum sensors (*1*) will eventually be able to transduce (*2*) quantum information in physical systems directly to a quantum memory (*3*, *4*), in which it can be processed by a quantum computer. Figure 1A illustrates the distinction between conventional and quantum-enhanced experiments. For example, in a quantum-enhanced experiment, multiple photons might be captured and stored coherently at each node of a quantum network and then processed coherently to extract an informative signal (*5*, *6*, *7*). In both the conventional and quantum-enhanced settings, multiple copies of the same quantum state are acquired. The crucial distinction is that the copies are measured one at a time in conventional experiments whereas entangling measurements across multiple copies are allowed in quantum-enhanced experiments.

[1]Institute for Quantum Information and Matter, Caltech, Pasadena, CA, USA. [2]Department of Computing and Mathematical Sciences, Caltech, Pasadena, CA, USA. [3]Google Quantum AI, Venice, CA 90291, USA. [4]Harvard Society of Fellows, Cambridge, MA 02138, USA. [5]Black Hole Initiative, Cambridge, MA 02138, USA. [6]Department of Electrical Engineering and Computer Science, University of California Berkeley, Berkeley, CA, USA. [7]Simons Institute for the Theory of Computing, Berkeley, CA, USA. [8]Microsoft Research AI, Redmond, WA 98052, USA. [9]Institute for Integrated Circuits, Johannes Kepler University Linz, Austria. [10]AWS Center for Quantum Computing, Pasadena, CA 91125, USA.
*Corresponding author. Email: hsinyuan@caltech.edu (H.-Y.H.); jmcclean@google.com (J.R.M.)

Recent mathematical analyses performed by some of the authors show that there exist properties of an n-qubit system that a quantum machine can learn efficiently whereas the requisite number of conventional experiments to achieve the same task is exponential in n (*8*, *9*). This exponential advantage contrasts sharply with the quadratic advantage achieved in many previously proposed strategies for improving sensing using quantum technology (*1*). In this article, we propose and analyze three classes of learning tasks with exponential quantum advantage and report on proof-of-principle experiments using up to 40 qubits on a Google Sycamore processor (*10*). These experiments confirm that a substantial quantum advantage can be realized even when the quantum memory and processor are both noisy.

To be more concrete, suppose that each experiment generates an n-qubit state ρ, and our goal is to learn some property of ρ (Fig. 1). We depict conventional and quantum-enhanced experiments for this scenario in Fig. 1B. In conventional experiments, each copy of ρ is measured separately, the measurement data are stored in a classical memory, and a classical computer outputs a prediction for the property after processing the classical data. In quantum-enhanced experiments, each copy of ρ is stored in a quantum memory, after which the quantum machine outputs the prediction after processing the quantum data in the quantum memory. We proved that for some tasks, the number of experiments needed to learn a desired property is exponential in n with the conventional strategy, but only polynomial in n using the quantum-enhanced strategy. For suitably defined tasks, we could achieve exponential quantum advantage using a protocol as simple as storing two copies of ρ in quantum memory and performing an entangling measurement. We also showed that quantum-enhanced experiments have a similar exponential advantage in a related scenario shown in Fig. 1C, in which the goal is to learn about a quantum process $\mathcal{E}$ rather than a quantum state ρ. Advantages of entangling measurements over single-copy measurements have been noticed previously (*11*, *12*), but our work goes much further by establishing an advantage that scales exponentially with system size.

Building on previous observations (*8*, *13*), we proved that for a task that entails acquiring information about a large number of noncommuting observables, quantum-enhanced experiments could have an exponential advantage even when the measured quantum state is unentangled. Our work substantially reduces the complexity of the required quantum-enhanced experiments, improving the prospects for near-term implementation. By performing experiments with up to 40 superconducting qubits, we showed that this quantum advantage persisted even when using currently available quantum processors. We also demonstrated quantum advantage in learning the symmetry class of a physical evolution operator, inspired by recent theoretical advances (*9*, *13*). Finally, in a theoretical contribution we rigorously proved that quantum-enhanced experiments have an exponential advantage in learning about the principal component of a noisy state, as previously indicated (*14*).

In our proof-of-principle experiments, we directly executed the state preparation or process to be learned within the quantum processor. In an actual application, the quantum data analyzed by the learning algorithm might be produced by an analog quantum simulator or a gate-based quantum computer. We also envision future applications in which quantum sensors equipped with quantum processors interact coherently with the physical world. The robustness of quantum advantage with respect to noise—validated by our experiments using a noisy superconducting device—boosts our confidence that the quantum-enhanced strategies described here can be exploited someday to achieve a substantial advantage in realistic applications.

Provable quantum advantage

We present three classes of learning tasks and the associated quantum-enhanced experiments, each yielding a provable exponential advantage over conventional experiments. Each result is encapsulated by a theorem which we state informally. Precise statements and proofs are presented in the supplementary materials. Our experimental demonstrations are discussed below in the section titled Demonstrations of Quantum Advantage. The proofs proceed by representing a classical algorithm with a decision tree depicted at the center of the gray robot in Fig. 1. The tree representation

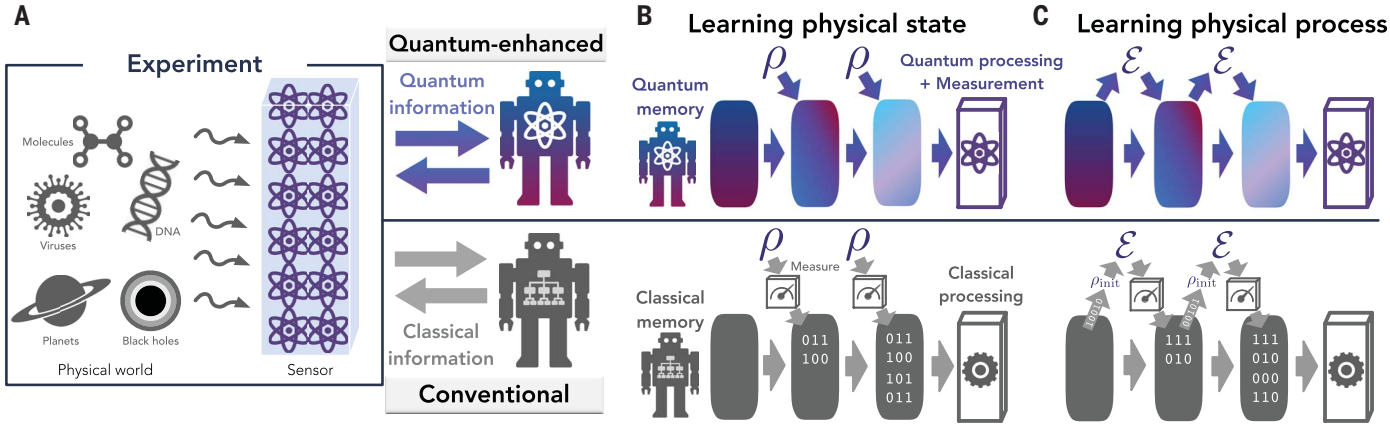


Fig. 1. Illustration of quantum-enhanced and conventional experiments. (**A**) Quantum-enhanced experiments versus conventional experiments. Quantum-enhanced or conventional experiments interface with a quantum or classical machine running a quantum or classical learning algorithm that can store and process quantum or classical information. (**B**) Learning physical state ρ. Each experiment produces a physical state ρ. In the conventional setting, we measure each ρ to obtain classical data (the measurement could depend on prior measurement outcomes) and store the data in a classical memory. In the quantum-enhanced setting, ρ can coherently alter the quantum information stored in the memory of the quantum machine (illustrated by the change in color). With large enough quantum memory, the quantum machine can simply store each copy of ρ. After multiple rounds of experiments, quantum processing followed by a measurement is performed on the quantum memory. (**C**) Learning physical process $\mathcal{E}$. Each experiment experiences evolution under $\mathcal{E}$. In the conventional setting, the classical machine specifies the input state to $\mathcal{E}$ by using a classical bitstring and obtains classical measurement data (*33*). In the quantum-enhanced setting, the evolution of $\mathcal{E}$ coherently alters the memory of the quantum machine: the input state to $\mathcal{E}$ is entangled with the quantum memory in the quantum machine and the output state is retrieved coherently by the quantum machine.

encodes how the classical memory changes as we obtain more experimental data. We then analyzed how the transitions on the tree differ for distinct measured physical systems to provide rigorous information-theoretic lower bounds. A general mathematical framework building on (*13*) is given in supplementary materials, section C.

The first task concerns learning about a physical system described by an n-qubit state, ρ. We suppose that each experiment generates one copy of ρ. In the conventional setting, we measure each copy of ρ to obtain classical data. The procedure can be adaptive, that is, each measurement can depend on the data obtained in earlier measurements. In the quantum-enhanced setting, a quantum computer can store each copy of ρ in a quantum memory and act jointly on multiple copies of ρ. In both scenarios we require all quantum data to be measured at the end of the learning phase of the procedure so that only classical data survives. After the learning is completed the learner is asked to provide an accurate prediction for the expectation value of one observable drawn from a set $\{O_1, O_2, ...\}$, where the number of observables in the set is exponentially large in n. The observables in the set can be highly incompatible, that is, each observable may fail to commute with many others in the set.

In prior work (*8*, *13*), we required the learner to predict exponentially many observables, which is not possible in practice if the system size is large. To demonstrate the advantage in an actual device, we proved that predicting just the absolute value of one observable requires exponentially many copies in the conventional scenario. By contrast, predicting the entire set of observables can be achieved with a polynomial number of copies in the quantum-enhanced scenario. We thereby established the following constant versus exponential separation. The proof is given in supplementary materials, section D.

Theorem 1 (Predicting observables): There exists a distribution over n-qubit states and a set of observables such that in the conventional scenario, at least order 2^n experiments are needed to predict the absolute value of one observable selected from the set, whereas a constant number of experiments suffice in the quantum-enhanced scenario.

The exponential quantum advantage can occur even if the state ρ is unentangled. For example, in our experiments we consider $\rho \propto (I + \alpha P)$, in which P is an n-qubit Pauli operator and $\alpha \in (-1, 1)$. This state can be realized as a probabilistic ensemble of product states, each of which is an eigenstate of P with eigenvalue α. Even if the state is known to be of this form but P and α are unknown, the exponential separation between conventional and quantum-enhanced experiments persists. Moreover, the quantum advantage can be achieved by performing simple entangling measurements on pairs of copies of ρ. That the quantum advantage applies even when correlations among the n qubits are classical leads us to believe that the quantum-enhanced strategy will be beneficial in a broad class of sensing applications. In supplementary materials section G we extend this theorem, showing that a sufficiently large quantum memory is needed to achieve this task in the quantum-enhanced scenario.

Our second ML task with a quantum advantage is quantum principal component analysis (PCA) (*14*). In this task each experiment produces one copy of ρ, and our goal is to predict properties of the (first) principal component of ρ, namely the eigenstate $|\psi\rangle$ of ρ with the largest eigenvalue. For example, we may want to predict the expectation values of a few observables in the state $|\psi\rangle$. This task may become a valuable component of future quantum-sensing applications. If an imperfect quantum sensor transduces a detected quantum state into quantum memory, the state is likely to be corrupted by noise. But it is reasonable to expect that properties of the principal component are relatively robust with respect to noise (*15*) and therefore highly informative about the uncorrupted state. To perform quantum PCA, a learning algorithm was introduced in (*14*) on the basis of phase estimation, which requires fault-tolerant quantum computers. One can also obtain information about the principal component of ρ by using more near-term algorithms, such as virtual cooling (*16*), virtual distillation (*17*, *18*), and variational algorithms (*19*, *20*).

Although the quantum PCA algorithm in (*14*) is exponentially faster than known algorithms

based on conventional experiments, this advantage was not proven against all possible algorithms in the conventional scenario. We rigorously established the exponential quantum advantage for performing quantum PCA. The exponential quantum advantage also holds in some of the near-term proposals (*16*, *17*). The proofs are provided in supplementary materials section E.

Theorem 2 (Performing quantum PCA): In the conventional scenario, at least order $2^{n/2}$ experiments are needed to learn a fixed property of the principal component of an unknown n-qubit quantum state, whereas a constant number of experiments will suffice in the quantum-enhanced scenario.

It is worth commenting on recent results in (*21*, *22*) showing that quantum PCA can be achieved by polynomial-time classical algorithms, which may seem to contradict Theorem 2. Those works assume the ability to access any entry of the exponentially large matrix ρ to exponentially high precision in polynomial time. Achieving such high precision requires measuring exponentially many copies of ρ, which takes an exponential number of experiments and exponential time. Hence, the assumptions of (*21*, *22*) do not hold here. See (*23*), which provides a detailed exposition of these matters.

Another core task in quantum mechanics is understanding physical processes rather than states. Here, each experiment implements a physical process $\mathcal{E}$, and we can interface with $\mathcal{E}$ through a quantum or classical machine in the quantum-enhanced or conventional setting; see Fig. 1C. We showed that a quantum machine can learn an approximate model of any polynomial-time quantum process $\mathcal{E}$ from only a polynomial number of experiments. Given a distribution on input states, the approximate model can predict the output state from $\mathcal{E}$ accurately on average. By contrast, we would need an exponential number of experiments to achieve the same task in the conventional setting. The proof for general quantum processes is given in supplementary materials, section F.

Theorem 3 (Learning quantum processes): Suppose we are given a polynomial-time physical process $\mathcal{E}$ acting on n qubits and a probability distribution over n-qubit input states. In the conventional scenario, at least order 2^n experiments are needed to learn an approximate model of $\mathcal{E}$ that predicts output states accurately on average, whereas a polynomial number of experiments will suffice in the quantum-enhanced scenario.

Demonstrations of quantum advantage

The exponential quantum advantage captured by Theorems 1, 2, and 3 applies no matter how much classical processing power is leveraged in the conventional experiments. The conventional strategy fails because there is simply no way to access enough classical data to perform the specified tasks if the number of experiments is subexponential in n. However, these exponential separations apply in an idealized setting in which quantum states are stored and processed perfectly. This leads us to ask whether access to quantum memory unlocks a substantial quantum advantage under more realistic conditions.

For two different tasks, we have investigated the robustness of the quantum advantage by conducting experiments with a superconducting quantum processor. We consider specialized tasks that maintain exponential quantum advantage and have better noise robustness than the general tasks described in the previous section. The first task we studied pertains to Theorem 1. The task is to approximately estimate the magnitude for the expectation value of Pauli observables. The unknown state is an unentangled n-qubit state $\rho = 2^{-n}(I + \alpha P)$, in which $\alpha = \pm 0.95$, P is a Pauli operator, and both α and P are unknown. After all measurements are completed and learning is terminated, two distinct Pauli operators, Q_1 and Q_2, are announced, one of which is P and the other of which is not equal to P. We then ask the machine to determine which of $|tr(Q_1\rho)|$ and $|tr(Q_2\rho)|$ is larger.

In the conventional scenario in which copies of ρ are measured one by one, the best known strategy is to use randomized Clifford measurements requiring an exponential number of copies to achieve the task with reasonable success probability (*8*, *24*). In the quantum-enhanced scenario, by contrast, copies of ρ are deposited in quantum memory two at a time and a Bell measurement across the two copies is performed to extract a snapshot of the state. In the quantum-enhanced scenario, we consider two different methods for analyzing the measurement data. The first method uses a specialized formula for estimating $|tr(Q\rho)|$, given in Appendix D2. Figure 2A depicts—as a function of the system size n—the number of experiments needed in the conventional and quantum-enhanced scenarios to achieve 70% prediction accuracy, in which the data from the quantum-enhanced experiments is analyzed by this first method. Also shown is a theoretical lower bound on the number of experiments needed in the conventional scenario, proven in Appendix D4. The first method is explicitly tailored to the structure of this particular learning problem and so cannot be applied readily to other problems. Our second method is more flexible and hence more broadly applicable; we make predictions by feeding the measurement data to a supervised ML model based on a recurrent neural network (*25*, *26*, *27*), as depicted in Fig. 2B. In contrast to the first method, the ML method does not require prior knowledge about the learning task. We train the neural network with noiseless simulation data for small system sizes ($n < 8$). We then use the neural network to make predictions when we are provided with experimental data for large system sizes $8 \le n \le 20$. We report the prediction accuracy, which is equal to the probability for correctly answering whether $|tr(Q_1\rho)|$ or $|tr(Q_2\rho)|$ is larger. Figure 2C shows the performance of the ML model as we train the neural network. Despite the noisy storage and processing in the experimental device, we observed a substantial quantum advantage using both the specialized and ML methods. Notably, when using ML, training on smaller systems sufficed for making good predictions on larger systems, a further indication that the measurement data in the quantum-enhanced scenario is so revealing that no special-purpose method is needed to extract a clear signal.

The second task we studied, which pertains to Theorem 3, was inspired by the recent observation that quantum-enhanced experiments can efficiently identify the symmetry class of a quantum evolution operator, whereas conventional experiments cannot (*9*, *13*). An unknown n-qubit quantum evolution operator is presented, drawn either from the class of all unitary transformations or the class of time-reversal-symmetric unitary transformations (i.e., real orthogonal transformations). We consider whether an unsupervised ML can learn to recognize the symmetry class of the unknown evolution operator on the basis of data obtained from either quantum-enhanced experiments or conventional experiments. An illustration is shown in Fig. 3A.

In the conventional scenario, we repeatedly apply the unknown evolution operator to the initial state $|0\rangle^{\otimes n}$ and then measure each qubit of the output state in the Y-basis. Under T-symmetric evolution the output state has purely real amplitudes; hence the expectation value of any purely imaginary observable, such as the Pauli-Y operator, is always zero. By contrast the expectation value of Y after general unitary evolution is generically nonzero but may be exponentially small and hence hard to distinguish from zero. In the quantum-enhanced scenario we make use of n additional memory qubits. We prepare an initial state in which the n system qubits are entangled with the n memory qubits, evolve the system qubits under the unknown evolution operator, swap the system and memory qubits, evolve the system qubits again, and finally perform n Bell measurements, each acting on one system qubit and one memory qubit.

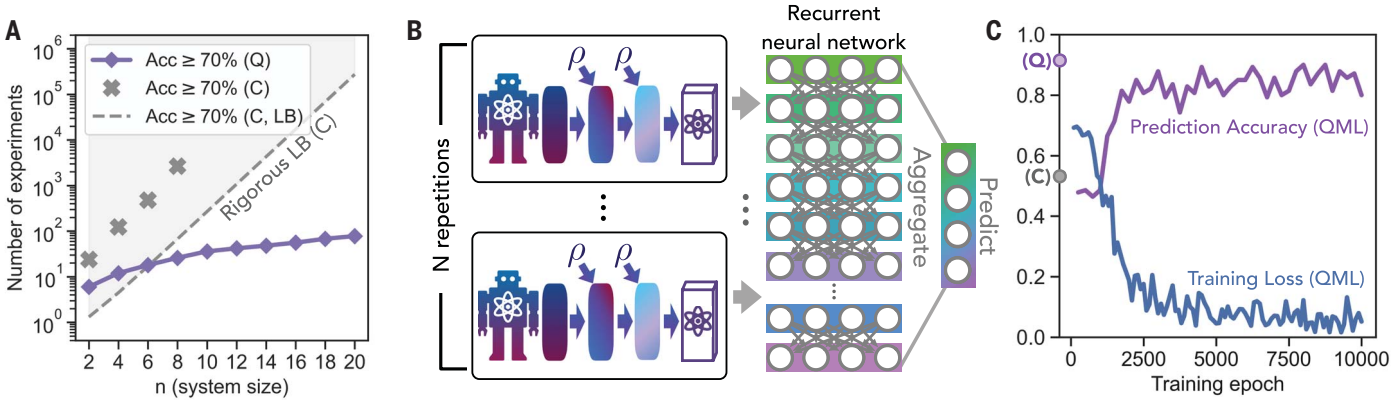


Fig. 2. Quantum advantage in learning physical states. (**A**) Quantum advantage in the number of experiments needed to achieve ≥70% accuracy. Here, Q corresponds to results running the best-known strategy for quantum-enhanced experiments, described in Appendix D2, and C corresponds to results running the best-known conventional strategy. The dotted line is a lower bound for any conventional strategy (C, LB) as proven in Appendix D4. Even running on a noisy quantum processor, quantum-enhanced experiments are seen to vastly outperform the best theoretically achievable conventional results (C, LB). (**B**) Supervised ML model based on quantum-enhanced experiments. *n* repetitions of quantum-enhanced experiments are performed and the data is fed into a gated recurrent neural network (GRU) (*25*, *26*). The neurons in the GRU are aggregated to predict an output. (**C**) Training process of the supervised ML model. We train the supervised ML model to determine which of two *n*-qubit Pauli operators has a larger magnitude for the expectation value in an unknown state ρ with noiseless simulation for small system sizes (n < 8). We consider the cross entropy (*34*) as the training loss. Then we use the supervised ML model to make predictions with data from noisy quantum-enhanced experiments running on the Sycamore processor (*10*) for larger system sizes ($8 \le n \le 20$). We consider the probability to predict correctly as the prediction accuracy. The purple (Q) and gray (C) dots on the y-axis are the accuracy of the best-known quantum-enhanced and conventional strategy considered in (A). Random guessing yields a prediction accuracy of 0.5.

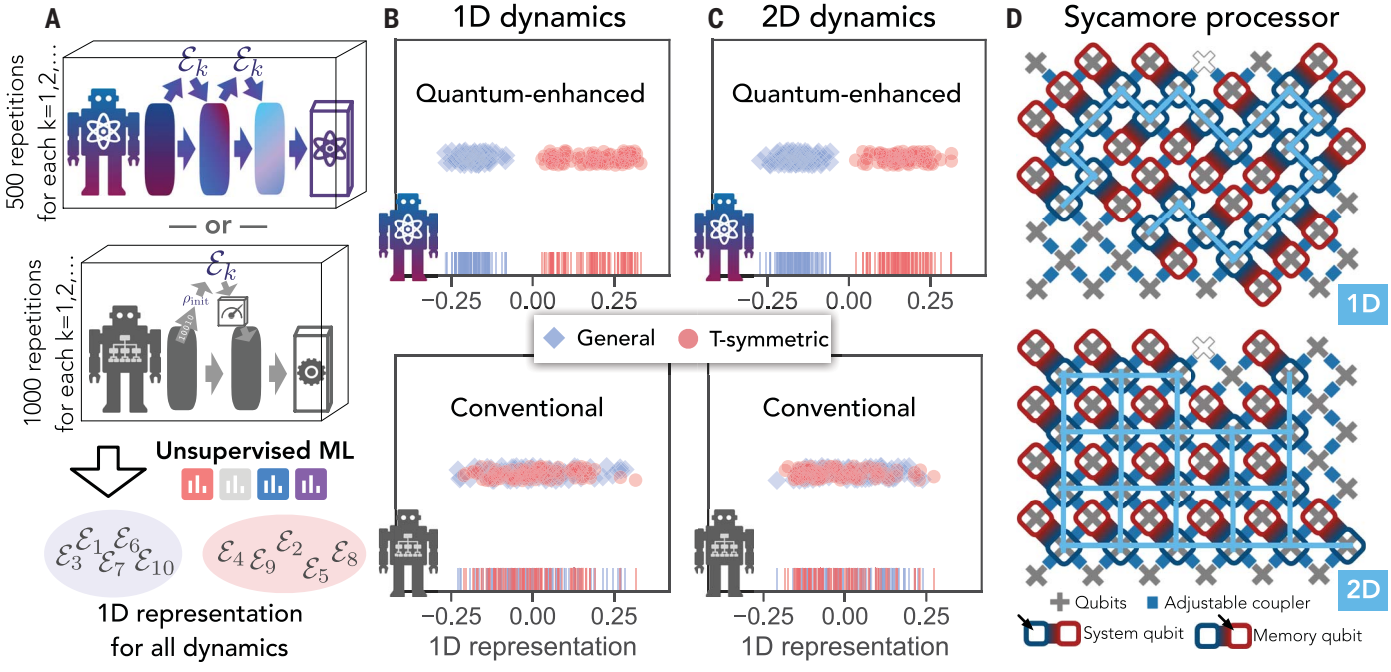


Fig. 3. Quantum advantage in learning physical dynamics. (**A**) Unsupervised MLmodel. We perform 500 repetitions of quantum-enhanced experiments (each accessing $\mathcal{E}_k$ twice) for every physical process $\mathcal{E}_k$ and feed the data into an unsupervised ML model (Gaussian kernel PCA) (*28*) to learn a 1D representation for describing distinct physical dynamics $\mathcal{E}_1, \mathcal{E}_2, \ldots$. Similarly, we also consider applying unsupervised ML to data obtained from 1000 repetitions of the best-known conventional experiments (each accessing $\mathcal{E}_k$ once) for every physical process $\mathcal{E}_k$. (**B**) Representation learned by unsupervised ML for 1D dynamics. Each point corresponds to a distinct physical process $\mathcal{E}_k$. The vertical line at the bottom shows the exact 1D representation of each $\mathcal{E}_k$. Half the processes satisfy time-reversal symmetry (blue diamonds) whereas the other half do not (red circles). When fed with data from quantum-enhanced experiments, the ML model accurately discovers the underlying symmetry pattern. By contrast, the ML model fails to do so when fed with data from conventional experiments. (**C**) Representation learned by unsupervised ML for 2D dynamics. (**D**) The geometry implemented on the Sycamore processor (*10*). We consider two different classes of connectivity geometry for implementing 1D (top) and 2D (bottom) dynamics.

Each evolution operator is a one-dimensional (1D) or 2D n-qubit quantum circuit as shown in Fig. 3D. After sampling many different evolution operators from both symmetry classes (and obtaining data from each sampled evolution multiple times), we used an unsupervised ML model (kernel PCA) (*28*) to find a 1D representation of the evolution operators. The representations learned by the unsupervised ML model are shown in Fig. 3, B and C. By using the quantum-enhanced data, the ML model discovers a clean separation between the two symmetry classes, whereas there is no discernable separation into classes when using data from conventional experiments. The signal from the quantum-enhanced experiments was strong enough that the two classes were easily recognized without access to any labeled training data.

In supplementary materials section A4, we analyzed the measurement data using the best-known special-purpose method specifically designed to distinguish general unitary transformations from real orthogonal transformations. We found a quantum advantage similar to that obtained with the ML model. The revelation that unsupervised learning yields results that are competitive with a more customized analysis highlights the potential for discovering previously unknown phenomena with quantum-enhanced measurement strategies. Properties that are blurred beyond recognition by single-copy measurements are brought into sharp relief by two-copy measurements.

Outlook

We have investigated how quantum technology can enhance our ability to discover unknown phenomena occurring in nature. For a variety of tasks, we proved that quantum-enhanced strategies that use quantum memory and quantum processing can predict properties of physical systems using exponentially fewer experiments than conventional strategies. This exponential advantage is achievable even if the amount of classical processing used in the conventional strategies is unlimited and when the physical system exhibits only classical correlations. Although many previous studies of quantum advantage have focused on computational tasks with known inputs, our work focused instead on learning tasks in which the goal is to learn about an a priori unknown physical system. This work provides a new approach to understanding and achieving quantum advantage in quantum ML (*29*,*30*) and quantum sensing (*1*).

Our experiments with up to 40 qubits in a superconducting quantum processor showed that a substantial quantum advantage is already evident when using today's noisy intermediate-scale quantum platforms (*31*). These experiments demonstrated that supervised and unsupervised ML models (*27*, *32*) employing data obtained from quantum-enhanced experiments could predict properties and discover underlying structure in physical systems that are beyond the scope of conventional experiments.

We envision that future quantum sensing systems will be able to transduce detected quantum data to a quantum memory and then process the stored data with a quantum computer. Although for now we lack suitably advanced sensors and transducers, we have conducted proof-of-concept experiments in which quantum data were directly planted in our quantum processor. Nevertheless, the robust quantum advantage we have validated highlights the potential for advancing quantum platforms to unlock facets of nature that would otherwise remain concealed.

REFERENCES AND NOTES

1. C. L. Degen, F. Reinhard, P. Cappellaro, *Rev. Mod. Phys.* **89**, 035002 (2017).
2. N. Lauk *et al*., *Quantum Sci. Technol.* **5**, 020501 (2020).
3. A. I. Lvovsky, B. C. Sanders, W. Tittel, *Nat. Photonics* **3**, 706–714 (2009).
4. E. Dennis, A. Kitaev, A. Landahl, J. Preskill, *J. Math. Phys.* **43**, 4452–4505 (2002).
5. D. Gottesman, T. Jennewein, S. Croke, *Phys. Rev. Lett.* **109**, 070503 (2012).
6. J. Bland-Hawthorn, M. J. Sellars, J. G. Bartholomew, *J. Opt. Soc. Am. B* **38**, A86–A98 (2021).
7. V. Giovannetti, S. Lloyd, L. Maccone, *Nat. Photonics* **5**, 222–229 (2011).
8. H.-Y. Huang, R. Kueng, J. Preskill, *Phys. Rev. Lett.* **126**, 190505 (2021).
9. D. Aharonov, J. Cotler, X.-L. Qi, Quantum algorithmic measurement. arXiv.2004.01372 [quant-ph] (2021)
10. F. Arute *et al*., *Nature* **574**, 505–510 (2019).
11. A. Peres, W. K. Wootters, *Phys. Rev. Lett.* **66**, 1119–1122 (1991).
12. C. H. Bennett *et al*., *Phys. Rev. A* **59**, 1070–1091 (1999).
13. S. Chen, J. Cotler, H.-Y. Huang, J. Li, in *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, 2022), pp. 574–585.
14. S. Lloyd, M. Mohseni, P. Rebentrost, *Nat. Phys.* **10**, 631–633 (2014).
15. B. Koczor, *New J. Phys.* **23**, 123047 (2021).
16. J. Cotler *et al*., *Phys. Rev. X* **9**, 031013 (2019).
17. W. J. Huggins *et al*., *Phys. Rev. X* **11**, 041036 (2021).
18. B. Koczor, *Phys. Rev. X* **11**, 031057 (2021).
19. R. LaRose, A. Tikku, É. O'Neel-Judy, L. Cincio, P. J. Coles, *npj Quantum Information* **5**, 1–10 (2019).
20. M. Cerezo, K. Sharma, A. Arrasmith, P. J. Coles, Variational quantum state eigensolver. arXiv.2004.01372 [quant-ph] (2020).
21. E. Tang, *Phys. Rev. Lett.* **127**, 060503 (2021).
22. N.-H. Chia *et al*., in *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing* (2020), pp. 387–400.
23. J. Cotler, H.-Y. Huang, J. R. McClean, Revisiting dequantization and quantum advantage in learning tasks. arXiv.2112.00811 [quant-ph] (2021).
24. H.-Y. Huang, R. Kueng, J. Preskill, *Nat. Phys.* **16**, 1050–1057 (2020).
25. J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, arXiv.1412.3555 [cs.NE] (2014).
26. D. Tang, B. Qin, T. Liu, in *Proceedings of the 2015 conference on empirical methods in natural language processing* (2015), pp. 1422–1432.
27. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning* (The MIT Press, 2016).
28. B. Schölkopf, A. Smola, K.-R. Müller, *Neural Comput.* **10**, 1299–1319 (1998).
29. J. Biamonte *et al*., *Nature* **549**, 195–202 (2017).
30. M. Broughton *et al*., Tensorflow quantum: A software framework for quantum machine learning. arXiv.2003.02989 [quant-ph] (2021).
31. J. Preskill, Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018).
32. M. Mohri, A. Rostamizadeh, A. Talwalkar, *Foundations of Machine Learning* (The MIT Press, 2018).
33. M. Mohseni, A. T. Rezakhani, D. A. Lidar, *Phys. Rev. A* **77**, 032322 (2008).
34. K. P. Murphy, *Machine Learning: a Probabilistic Perspective* (MIT press, 2012).
35. H.-Y. Huang *et al*., Code for quantum advantage in learning from experiments, Github (2022); https://github.com/quantumlib/ReCirq/tree/master/recirq/qml_lfe.
36. H.-Y. Huang *et al*., *Zenodo* (2022); https://zenodo.org/record/6400225#.YoZ23ajMI2w.

ACKNOWLEDGMENTS

The quantum hardware used for this experiment was developed by the Google Quantum AI hardware team, under the direction of A. Megrant, J. Kelly, and Y. Chen. Methods for device calibrations were developed by the physics team led by V. Smelyanskiy. Data were collected via cloud access through Google's Quantum Computing Service. We thank B. Foxen for special support and maintaining the device to the caliber needed to complete the experiments. **Funding:** H.H. is supported by a Google PhD Fellowship. J.C. is supported by a Junior Fellowship from the Harvard Society of Fellows, by the Black Hole Initiative, and in part by the Department of Energy under grant DE-SC0007870. S.C. is supported by the National Science Foundation under Award 2103300 and was visiting the Simons Institute for the Theory of Computing while part of this work was completed. J.P. acknowledges funding from the US Department of Energy Office of Science Office of Advanced Scientific Computing Research (DE-NA0003525, DE-SC0020290), and the National Science Foundation (PHY-1733907). The Institute for Quantum Information and Matter is an NSF Physics Frontiers Center. **Author contributions:** H.H., J.C., S.C., J.L., R.K., J.P., and J.M. were involved in conceptualization, planning, and theoretical developments. H.H., M.B., M.M., H.N., R.B., and J.M. contributed to the design and execution of the experiments on the Google processor. All authors were involved in the writing and presentation of the work. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** In addition to the data in the paper and supplemental materials, code related to this experiment is hosted at Github (*35*). The data needed to reproduce figures are hosted at Zenodo (*36*). All other data needed to evaluate the conclusions in the paper are present in the paper or the 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.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn7293
Materials and Methods
Supplementary Text
Figs. S1 to S15
Table S1
Appendices A to G
References (*37*–*68*)

Submitted 17 December 2021; accepted 14 April 2022
10.1126/science.abn7293

PLANT SCIENCE

Organic acids and glucose prime late-stage fungal biotrophy in maize

Matthias Kretschmer[1], Djihane Damoo[1], Sherry Sun[1]†, Christopher W. J. Lee[1], Daniel Croll[2], Harry Brumer[3], James Kronstad[1]*

Many plant-associated fungi are obligate biotrophs that depend on living hosts to proliferate. However, little is known about the molecular basis of the biotrophic lifestyle, despite the impact of fungi on the environment and food security. In this work, we show that combinations of organic acids and glucose trigger phenotypes that are associated with the late stage of biotrophy for the maize pathogen *Ustilago maydis*. These phenotypes include the expression of a set of effectors normally observed only during biotrophic development, as well as the formation of melanin associated with sporulation in plant tumors. *U. maydis* and other hemibiotrophic fungi also respond to a combination of carbon sources with enhanced proliferation. Thus, the response to combinations of nutrients from the host may be a conserved feature of fungal biotrophy.

Fungi threaten human health, crop production, and food security (*1*, *2*). Many economically important fungal pathogens of plants are obligate biotrophs that cannot be propagated outside of the host (*3*). Obligate biotrophs also include beneficial mycorrhizal fungi that provide critical nutrients such as phosphate to 80% of plant species (*4*). In general, there is a lack of information about the nutritional requirements for fungal proliferation and development in host tissue, although genome analyses suggest that the loss of specific biosynthetic capabilities conditions a reliance on host nutrients (*3*, *5*).

The maize fungal pathogen *Ustilago maydis* can be grown axenically in culture but is obligately dependent on a plant host to complete the sexual phase of its life cycle (*6*–*8*). This phase involves the mating of compatible sporidia to establish invasive filaments; the delivery of effector proteins; the induction of conspicuous tumors on leaves, ears, and tassels; and the formation of massive numbers of melanized spores in tumors (*8*). In addition to the economic impact of *U. maydis* on maize production, the fungus is unusual because the tumor tissue has been prized as a culinary delicacy in Mexico since the time of the Aztecs (*9*).

Induction of biotrophic phenotypes in culture

During infection, *U. maydis* reprograms developing tumor tissue into a sink for photosynthate, and the carbon sources available to the fungus include carbohydrates and organic

[1]Michael Smith Laboratories and Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada. [2]Laboratory of Evolutionary Genetics, Institute of Biology, Université de Neuchâtel, Neuchâtel, Switzerland. [3]Michael Smith Laboratories and Department of Chemistry, University of British Columbia, Vancouver, BC, Canada.
*Corresponding author. Email: kronstad@msl.ubc.ca
†Present address: Department of Botany, University of British Columbia, Vancouver, BC, Canada.

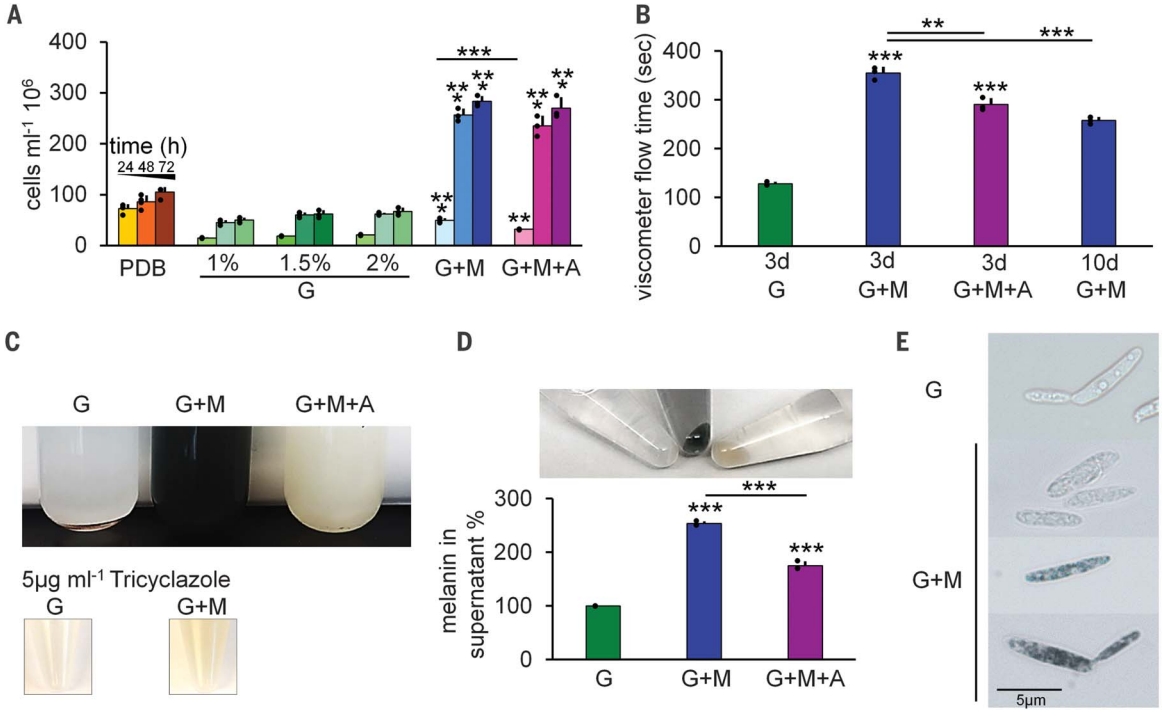


Fig. 1. Carbon sources influence proliferation, extracellular polysaccharide, and melanin. (**A**) Cell numbers were compared upon growth in minimal medium with different concentrations of glucose (G), glucose plus malate (G+M), glucose plus malate with aeration (G+M+A), or potato dextrose broth (PDB). The G+M and G+M+A conditions were compared with the G 1.5% condition. (**B**) Culture viscosity to detect extracellular polysaccharide was measured after 3 or 10 days of growth with a viscometer and is reported in seconds of flow time. (**C**) Detection of melanin in *U. maydis* cultures in G, G+M, or G+M+A after 72 hours of growth. Melanin formation in the G+M medium is inhibited with 5 µg ml^{-1} of tricyclazole. (**D**) Melanin is associated with cell pellets and is measurable in culture supernatants. (**E**) Melanin is cell associated with a range of intensities. In (A), (B), and (D), lines above the bar graphs indicate the statistical significance for the comparisons at the ends of each line. Significance levels for comparisons of the wild-type strain in different media are $^{**}P \leq 0.01$ and $^{***}P \leq 0.001$ according to analysis of variance (ANOVA) with a Tukey procedure as a post hoc test. Error bars indicate SD.

acids (*7*, *10*–*14*). Given that maize plants carry out C4 photosynthesis of the nicotinamide adenine dinucleotide phosphate (NADP)–malic enzyme subtype in which 75% of CO_2 is initially fixed into malate, we hypothesized that metabolic adaptation to organic acids is a key determinant of biotrophic proliferation for *U. maydis* (*7*, *15*). We tested this hypothesis by culturing the fungus in standard glucose medium with the addition of malate [glucose plus malate (G+M)] and found that this combination stimulated cell proliferation, increased culture viscosity, and triggered the accumulation of dark, pigmented cells (Fig. 1 and figs. S1 and S2). Other organic acids also triggered the same phenotypes in combination with glucose (fig. S3 and table S1). The increase in viscosity was due to the accumulation of extracellular polysaccharides with a β-1,3 glucan structure commonly found in fungi (fig. S2). The pigment was cell associated and was related to melanin, as determined with the specific inhibitor tricyclazole (Fig. 1, C to E, and fig. S1). These phenotypic changes prompted an investigation of the relevance of the observed responses to the biotrophic development of *U. maydis* in maize. We therefore examined melanin formation, the role of organic acid transporters, the transcription of genes for biotrophic effectors, and the contributions of mitochondrial functions and oxygen.

Melanin formation during *U. maydis* sporulation in tumors is catalyzed by the laccase Lac1 and the polyketide synthase Pks1 (*16*). However, additional enzymes may contribute to melanin formation because the genome encodes five other candidate laccases and four additional polyketide synthases (*17*, *18*). RNA sequencing analysis of cells grown in glucose (G) versus cells from the G+M condition revealed that the transcripts for three *pks* genes (*pks3*, *pks4*, and *pks5*) were elevated in the G+M condition (Fig. 2A, figs. S4 and S5, and tables S2 and S3). These *pks* genes are present in a cluster of 15 genes on chromosome 12, and their transcript levels are regulated by the transcription factor Mtf1 encoded within the cluster (Fig. 2A) (*18*). Ten genes in this cluster, including *mtf1* (up-regulated 1714-fold), are expressed at a late stage of infection [12 days post-inoculation (dpi)] (*14*). The transcripts for *mtf1* were elevated in the G+M condition, and we found that an *mtf1* deletion mutant did not

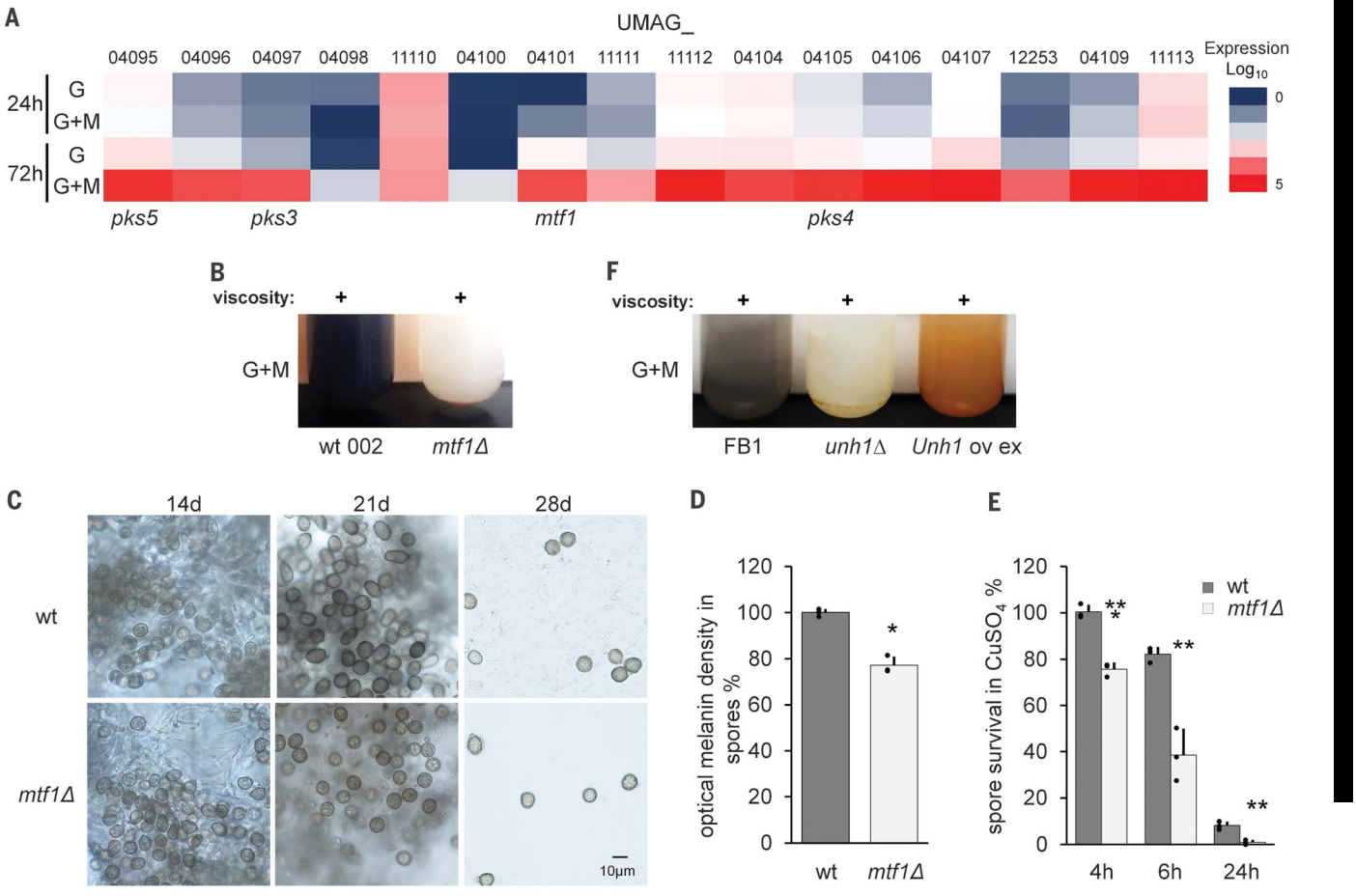


Fig. 2. Regulation and contributions of a melanin gene cluster. (**A**) A gene cluster encodes the transcriptional regulator Mtf1 and three polyketide synthases. The heatmap shows averaged normalized expression values of three biological replicates for G and G+M at 24 and 72 hours. The expression values were $\log_{10}$ transformed; the color code for the $\log_{10}$ scale is from 0 (low expression; dark blue) to 5 (high expression; red). Sample expression values between 0 and 1 were set to 1 before $\log_{10}$ transformation. (**B**) Melanin is reduced in the *mtf1*Δ mutant upon growth in G+M medium. wt, wild type. (**C**) Spores of the *mtf1*Δ mutant have reduced melanin formation over time during infection of maize seedlings. (**D**) Optical measurement of melanin revealed reduced content in spores of the *mtf1*Δ mutant at 28 days. (**E**) Survival of *mtf1*Δ spores is reduced upon treatment with $CuSO_4$. (**F**) Deletion or overexpression of the sporulation transcription factor *unh1* influences melanin formation. In (D) and (E), significance levels for comparisons of mutant and wild-type strains are $^*P \le 0.05$, $^{**}P \le 0.01$, and $^{***}P \le 0.001$ according to *t* test or ANOVA with a Tukey procedure as a post hoc test. Error bars indicate SD.

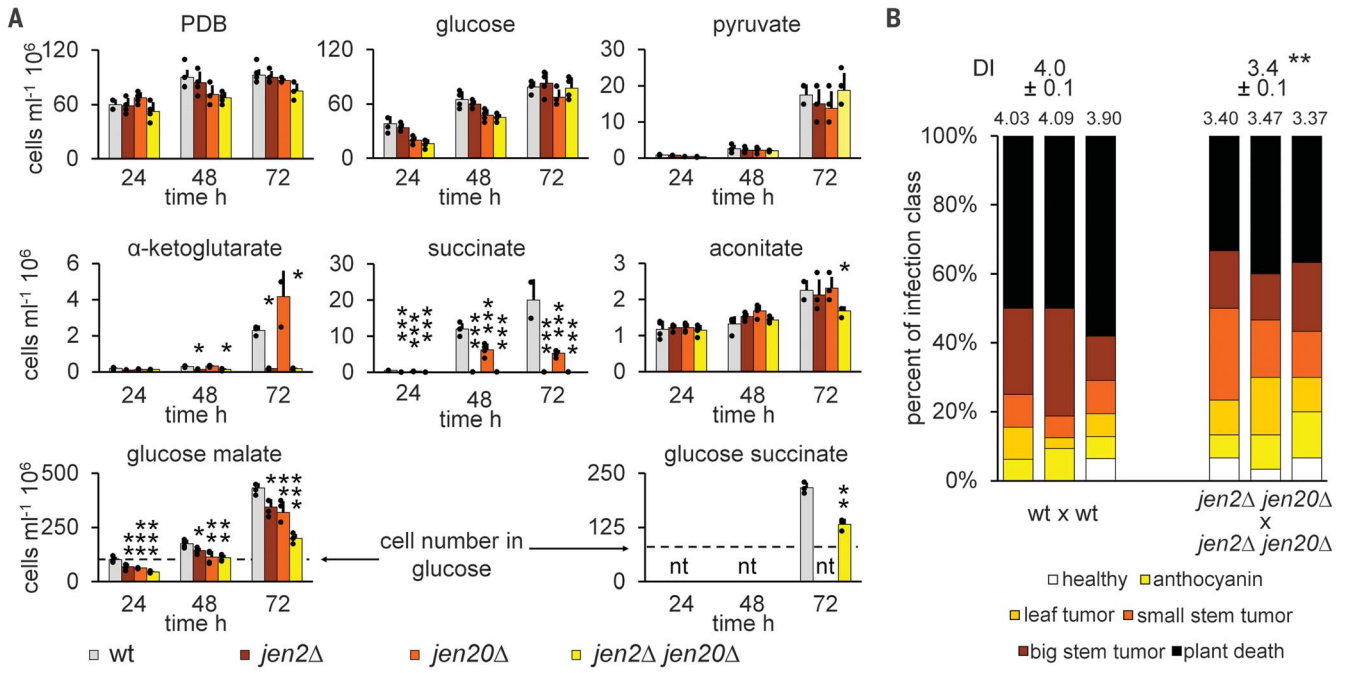


Fig. 3. Candidate transporters influence growth on dicarboxylates and virulence. (**A**) The *jen2*Δ and *jen20*Δ mutants have impaired growth on carboxylates and a combination of glucose and dicarboxylates. (**B**) Virulence of the *jen2*Δ *jen20*Δ double mutant is reduced on maize seedlings. Significance levels for comparisons of mutant and wild-type strains are $*P \leq 0.05$, $**P \leq 0.01$, and $***P \leq 0.001$ according to a *t* test for infection or ANOVA with a Tukey procedure as a post hoc test. A Kruskal-Wallis with Dunn analysis as a post hoc test was used to evaluate growth on aconitate. Error bars indicate SD. DI, disease index; nt, not tested.

form melanin during growth in G+M medium. Furthermore, the mutant showed reduced melanin content in spores from tumor tissue (Fig. 2, C and D). That is, the mutant caused disease in maize seedlings, which led to spore development, but the melanin content of isolated *mtf1*Δ spores was reduced by 22.8% (Fig. 2D). Additionally, the *mtf1*Δ spores showed incomplete maturation because their survival was reduced upon $CuSO_4$ treatment compared with wild-type spores (Fig. 2E).

The transcription factor Unh1 also regulates sporulation in *U. maydis*, and *unh1* transcripts were elevated during growth on G and G+M media at 72 hours versus G at 24 hours, as well as in infected plants (table S3) (*17*, *19*). The *unh1*Δ mutant was also compromised for melanin formation on G+M, thus further linking in planta sporulation to the melanin phenotype induced in culture (Fig. 2F). Complementation of the *unh1*Δ mutation resulted in a modified pigment color, perhaps due to overexpression of the gene (Fig. 2F) (*19*). In addition to Unh1, several other virulence-associated regulatory factors, including protein kinases and transcription factors, also influenced melanin formation in the G+M condition (fig. S6). Therefore, these functions are candidate components of a regulatory network for melanin formation. Overall, these results suggest that *U. maydis* has one melanin biosynthesis pathway that is dependent on Lac1 and Pks1 and a second pathway that is regulated by Mtf1 and Unh1; the latter pathway is induced during sporulation in planta and in response to growth in glucose plus organic acids.

The relevance of the response to organic acids for biotrophy was tested further by constructing mutants that lacked dicarboxylate transporters and examining virulence in maize seedlings. We mined the genome to identify candidate transporters and examined their transcript levels in cells from the G+M condition (tables S2, S4, and S5). From this analysis, we identified two dicarboxylate transporters, Jen2 and Jen20, that were required for robust growth on specific organic acids (e.g., aconitate, α-ketoglutarate, succinate, or malate) and in combination with glucose (Fig. 3A). Deletion of both *jen2* and *jen20* attenuated virulence on maize, although some disease symptoms still occurred, indicating that additional transporters contribute to in planta growth (Fig. 3B). Overall, these results indicate that the ability to acquire organic acids contributes to the virulence of *U. maydis* on maize.

In vitro expression of disease effectors

The delivery of effector proteins to suppress plant defense and promote virulence is a key aspect of biotrophic development (*14*). For *U. maydis*, candidate effectors are expressed in transcriptional modules of coexpressed genes at defined stages of infection, including plant surface interactions, establishment of biotrophy, nutrient acquisition, and induction of tumors (*14*, *20*). Many of the effectors are expressed only during growth in the host and define virulence-specific modules (*14*). We compared the transcriptional response to the G and G+M conditions with the established modules and found that transcripts encoding a subset of effectors involved in biotrophic development were elevated in response to carbon sources (figs. S7 to S9 and table S2). That is, the transcript levels for some effectors were more highly elevated in G+M at 72 hours than in either G at 24 hours or G at 72 hours, as was demonstrated for the specific effectors Eff1-1, Ten1, Rsp3, Afu1-3, Mig2-2, and Rrm67 (fig. S8) (*20*). Some of the effector genes also displayed elevated transcripts in G at 72 hours versus G at 24 hours, indicating a response to glucose depletion during the stationary phase. These effectors included the Afu3, See1, Pit2, Cmu1, Stp1, Stp4, Mig2-4, and Eff1 proteins (fig. S8 and table S2) (*20*, *21*). Many of the effector genes in *U. maydis* are found in clusters that affect virulence upon deletion (fig. S9) (*7*). In this regard, transcripts for genes in the major virulence clusters 2A, 6A, 10A, and 19A were elevated in the G+M at 72 hours condition

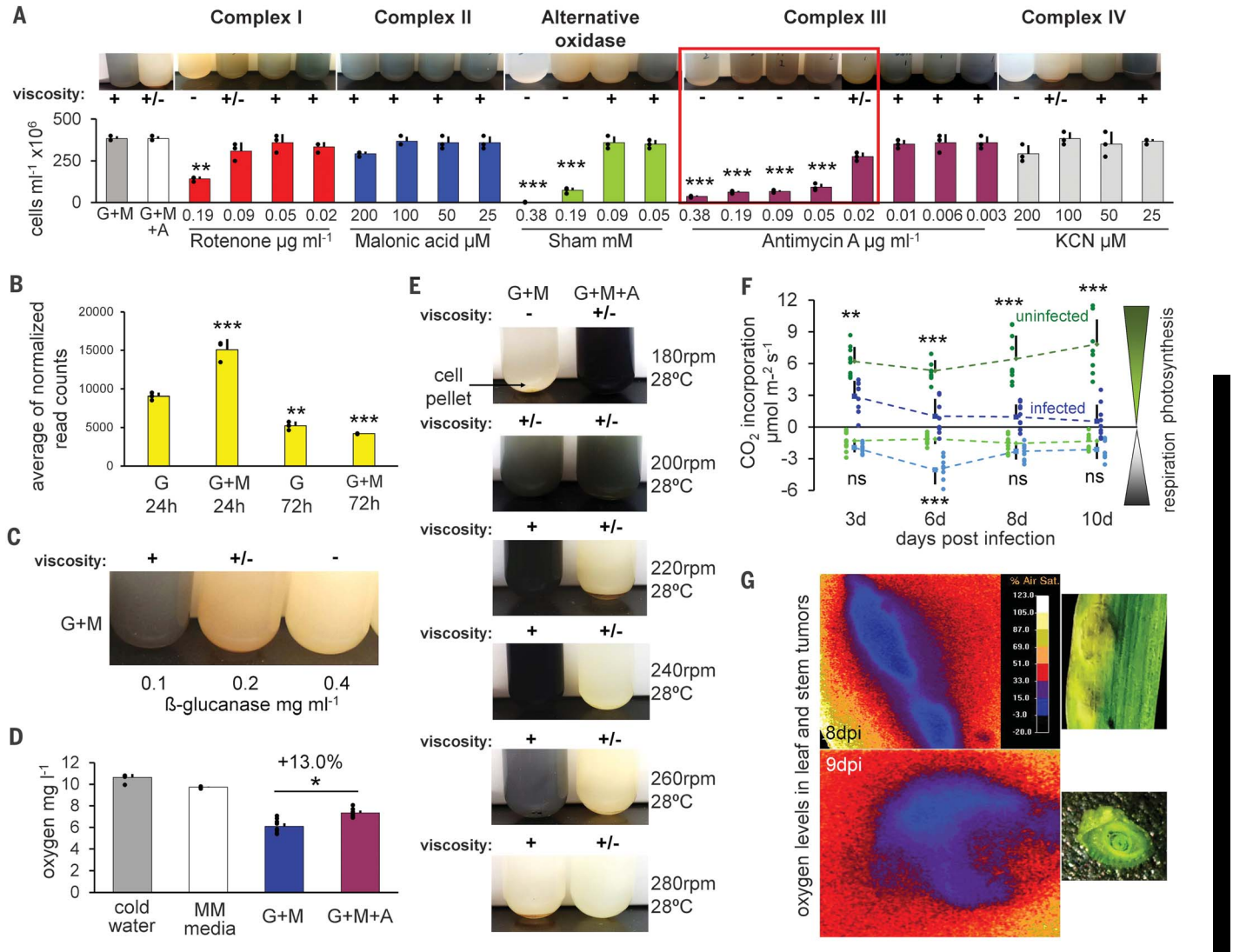


Fig. 4. Mitochondrial functions and oxygen contribute to biotrophic phenotypes. (**A**) Inhibition of electron transport chain complexes influences growth, culture viscosity, and melanin status in G+M. The red frame indicates melanin formation in cultures with growth inhibition. (**B**) Analysis of the normalized transcript levels of 155 genes with elevated transcript levels in the yellow general growth module (*14*) in G+M at 24 hours. Of these genes, 49 encode mitochondrial proteins. The expression of Fe-S cluster and electron transport chain genes can be found fig. S11 and table S2. (**C**) Influence of β-glucanase treatment on viscosity and melanin formation. (**D**) Oxygen levels in G+M cultures at 72 hours were measured and compared with those in cold water, uninoculated minimal medium (MM), and G+M+A cultures. (**E**) Melanin formation and culture viscosity are dependent on aeration as influenced by shaker speed. (**F**) Photosynthesis and respiration rates of control maize leaves (green) and infected leaves (blue) during a time course of infection. The dashed lines indicate trends over time. ns, not significant. (**G**) Two-dimensional depiction of oxygen levels in planta during infection at 8 and 9 dpi in leaves and stems with tumors. Yellow and orange, high oxygen levels; blue, low oxygen levels. Significance levels for comparisons between treated and untreated cells (A), between G and G+M (B), and between infected and uninfected tissue at each time point (F) are $*P \leq 0.05$, $**P \leq 0.01$, and $***P \leq 0.001$ according to a *t* test for oxygen measurement or ANOVA with a Tukey procedure as a post hoc test. Error bars indicate SD.

versus in G at 72 hours (fig. S9 and table S2). Transcript levels for genes in other clusters were also elevated in the G+M conditions (fig. S10). These clusters contained genes for the biosynthesis of itaconic acid and the siderophore ferrichrome A that are highly expressed in vivo (*22*), as well as genes for melanin and primary or secondary metabolism (fig. S10 and table S3). Overall, we conclude that the G+M culture condition triggers the transcription of genes normally expressed only during biotrophic growth in planta.

Mitochondrial functions and oxygen influence biotrophy

Mitochondrial functions and oxygen sensing play roles in the metabolic adaptation of fungi to the host environment (*23*). We found that mitochondrial functions (e.g., electron transport chain components and Fe-S–requiring proteins) were enriched in the Functional Catalogue (FunCat) analysis of the transcriptome data (figs. S4 and S11 and tables S2 and S6). We identified 155 genes in the module of general growth (*14*) that had elevated transcripts at 24 hours in G+M versus G; one-third of these genes are annotated as encoding mitochondrial

proteins. Transcripts related to Fe-S cluster–containing proteins of the electron transport chain complexes I to III, as well as transcripts for other components of complexes I to III and alternative oxidase, were elevated in the G+M condition (Fig. 4, A and B, and tables S2 and S6). We therefore tested the influence of inhibition of electron transport chain complexes on the response to mixed carbon sources and found that inhibitors of complex I, the alternative oxidase, and complex IV reduced melanin and culture viscosity but allowed growth in the G+M condition (Fig. 4A). In particular, inhibition of complex III reduced growth and viscosity, but melanin formation was still observed, perhaps indicating an uncoupling of pigmentation and proliferation.

Oxygen is the terminal electron acceptor in the electron transport chain, so we examined conditions with enhanced aeration (A) to assess the role of oxygen in melanin formation (G+M+A) (Fig. 4, C to E). Specifically, aeration was enhanced in the G+M condition by the addition of β-glucanase to reduce viscosity, which led to a concentration-dependent reduction in melanin (Fig. 4C). The G+M+A cultures resulted in 13% higher oxygen levels than the G+M condition (Fig. 4D), and the speed of culture shaking influenced melanin formation, as expected for a negative influence of oxygen (Fig. 4E and fig. S1). It is possible that oxygen could affect not only mitochondrial functions but also oxidation reactions that influence the polymerization of melanin precursors. Oxygen levels may also be relevant during pathogenic development in planta, especially given that infection negatively influences transcript levels for chloroplast genes, including those encoding photosynthetic functions (*24*, *25*). We evaluated the rates of photosynthesis and respiration and found that photosynthesis was inhibited in infected tissue, especially during tumor formation, whereas respiration peaked at 6 dpi (Fig. 4F). The observed down-regulation of photosynthesis is consistent with the results of previous studies (*25*). Noninvasive two-dimensional oxygen measurements confirmed that oxygen levels were also reduced in tumor tissue (Fig. 4G). We conclude that reduced oxygen tension is required for the in vitro biotrophic response to organic acids. Along with the impact of electron transport chain inhibition, this result implicates the mitochondria and mitochondrial functions as potential regulators of fungal biotrophy, perhaps through an influence on metabolite concentrations and the generation of signals that regulate gene expression.

We also tested additional fungal species for enhanced proliferation, viscosity, or pigmentation in media with glucose and organic acids to determine whether the response might be a general feature of biotrophic and hemibiotrophic fungi (fig. S12). We did not see an influence on the human pathogens *Candida albicans* and *Cryptococcus neoformans*, the saprophyte *Saccharomyces cerevisiae*, the mycorrhizal fungus *Laccaria bicolor*, or the necrotrophic plant pathogen *Sclerotinia sclerotiorum*. By contrast, the biotrophic and hemibiotrophic fungi *Ustilago hordei*, *Sporisorium reilianum*, *Fusarium oxysporum*, and *Verticillium dahliae* showed increased cell numbers or growth rates on the G+M medium compared with on G alone (fig. S12). Therefore, the response to mixed carbon sources may be a conserved feature of some biotrophic and hemibiotrophic fungi.

This study reveals a complex response of *U. maydis* to organic acids that involves mitochondrial functions, oxygen sensing, specific transporters, and transcriptional regulators of traits related to biotrophy. We also observed an accumulation of β-glucan in our culture conditions. This extracellular polysaccharide is a likely component of the mucilage matrix that accumulates during sporulation in tumors (*26*). The response to combinations of carbon sources may be a conserved feature of biotrophic fungal pathogens as well as other microbes that associate with plants. For example, obligate mycorrhizal fungi respond to lipids and sugars with proliferation and pre-spore formation, and rhizobia bacteria respond to dicarboxylates such as succinate and malate from legume hosts during nodulation and nitrogen fixation (*27*–*31*). Furthermore, features of bacterial nodulation such as extracellular polysaccharide production, malate metabolism, and oxygen sensing are shared with fungal tumor formation (*32*, *33*). The response to defined combinations of nutrients may therefore be a general theme in plant-microbe interactions.

REFERENCES AND NOTES

1. M. C. Fisher *et al*., *mBio* **11**, e00449-20 (2020).
2. J. B. Ristaino *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2022239118 (2021).
3. P. D. Spanu, R. Panstruga, *Front. Plant Sci.* **8**, 192 (2017).
4. N. Ferrol, C. Azcón-Aguilar, J. Pérez-Tienda, *Plant Sci.* **280**, 441–447 (2019).
5. P. D. Spanu *et al*., *Science* **330**, 1543–1546 (2010).
6. J. J. Christensen, "Corn smut caused by *Ustilago maydis*" (American Phytopathology Society Monograph no. 2, American Phytopathology Society, 1963).
7. J. Kämper *et al*., *Nature* **444**, 97–101 (2006).
8. W. Zuo *et al*., *Annu. Rev. Phytopathol.* **57**, 411–430 (2019).
9. M. Juárez-Montiel, S. Ruiloba de León, G. Chávez-Camarillo, C. Hernández-Rodríguez, L. Villa-Tanaca, *Rev. Iberoam. Micol.* **28**, 69–73 (2011).
10. R. J. Horst, T. Engelsdorf, U. Sonnewald, L. M. Voll, *J. Plant Physiol.* **165**, 19–28 (2008).
11. R. J. Horst *et al*., *Plant Physiol.* **152**, 293–308 (2010).
12. D. Sosso *et al*., *Plant Physiol.* **179**, 1373–1385 (2019).
13. M. Ludwig, *Front. Plant Sci.* **7**, 647 (2016).
14. D. Lanver *et al*., *Plant Cell* **30**, 300–323 (2018).
15. M. D. Hatch, *Biochem. J.* **125**, 425–432 (1971).
16. E. Islamovic *et al*., *Mol. Plant Microbe Interact.* **28**, 42–54 (2015).
17. M. Tollot *et al*., *PLOS Pathog.* **12**, e1005697 (2016).
18. E. Z. Reyes-Fernández, Y. M. Shi, P. Grün, H. B. Bode, M. Bölker, *Appl. Environ. Microbiol.* **87**, e01510-20 (2021).
19. C. E. Doyle, H. Y. Kitty Cheung, K. L. Spence, B. J. Saville, *Fungal Genet. Biol.* **94**, 54–68 (2016).
20. D. Lanver *et al*., *Nat. Rev. Microbiol.* **15**, 409–421 (2017).
21. M. Schuster, G. Schweizer, R. Kahmann, *Fungal Genet. Biol.* **112**, 21–30 (2018).
22. M. Kretschmer, D. Croll, J. W. Kronstad, *Mol. Plant Pathol.* **18**, 1222–1237 (2017).
23. B. Black, C. Lee, L. C. Horianopoulos, W. H. Jung, J. W. Kronstad, *PLOS Pathog.* **17**, e1009661 (2021).
24. M. Kretschmer, D. Croll, J. W. Kronstad, *Mol. Plant Pathol.* **18**, 1210–1221 (2017).
25. G. Doehlemann *et al*., *Plant J.* **56**, 181–195 (2008).
26. F. Banuett, I. Herskowitz, *Development* **122**, 2965–2976 (1996).
27. L. H. Luginbuehl *et al*., *Science* **356**, 1175–1178 (2017).
28. Y. Jiang *et al*., *Science* **356**, 1172–1175 (2017).
29. A. Keymer *et al*., *eLife* **6**, e29107 (2017).
30. Y. Sugiura *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 25779–25788 (2020).
31. C. C. M. Schulte *et al*., *Sci. Adv.* **7**, eabh2433 (2021).
32. S. Acosta-Jurado, F. Fuentes-Romero, J. E. Ruiz-Sainz, M. Janczarek, J. M. Vinardell, *Int. J. Mol. Sci.* **22**, 6233 (2021).
33. A. L. Le Gac, T. Laux, *Mol. Plant* **12**, 1422–1424 (2019).

ACKNOWLEDGMENTS

We thank G. Bakkeren, F. Banuett, G. Braus, C. Haney, K. Heimel, R. Kahmann, M. Perlin, and B. Saville for fungal strains and K. Heimel for insightful comments. We also thank the Department of Civil Engineering at UBC for the use of the oxygen meter and the Department of Botany for the use of the photosynthesis meter. **Funding:** This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada and by an NSERC-CREATE program (to J.K.) and by the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, US Department of Energy, grant (DE-SC0015662) to Parastoo Azadi at the Complex Carbohydrate Research Center. J.K. is a Burroughs Wellcome Fund Scholar in Molecular Pathogenic Mycology and a fellow of the CIFAR program: Fungal Kingdom Threats & Opportunities. **Author contributions:** Conceptualization: M.K. and J.K.; Methodology: M.K., D.D., S.S., and H.B.; Investigation: M.K., D.D., S.S., C.W.J.L., D.C., and H.B.; Visualization: M.K. and D.C.; Funding acquisition: J.K.; Project administration: J.K.; Supervision: J.K. and M.K.; Writing – original draft: M.K., J.K.; Writing – review and editing: M.K., J.K., D.D., S.S., C.W.J.L., D.C., and H.B. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The RNA sequencing data for the comparisons of transcript levels for cells grown in G medium versus G+M (with and without additional aeration) are available at the Gene Expression Omnibus with the accession number GSE194157 for the project and accession numbers GSM5829810 to GSM5829827 for the samples. All other data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials, and all materials are available from the authors without restriction. **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.abo2401
Materials and Methods
Figs. S1 to S12
Tables S1 to S8
References (*34*–*63*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 21 January 2022; accepted 4 May 2022
10.1126/science.abo2401

AGING

Circadian alignment of early onset caloric restriction promotes longevity in male C57BL/6J mice

Victoria Acosta-Rodríguez[1], Filipa Rijo-Ferreira[1,2]†, Mariko Izumo[1], Pin Xu[1], Mary Wight-Carter[3], Carla B. Green[1]*, Joseph S. Takahashi[1,2]*

Caloric restriction (CR) prolongs life span, yet the mechanisms by which it does so remain poorly understood. Under CR, mice self-impose chronic cycles of 2-hour feeding and 22-hour fasting, raising the question of if it is calories, fasting, or time of day that is the cause of this increased life span. We show here that 30% CR was sufficient to extend the life span by 10%; however, a daily fasting interval and circadian alignment of feeding acted together to extend life span by 35% in male C57BL/6J mice. These effects were independent of body weight. Aging induced widespread increases in gene expression associated with inflammation and decreases in the expression of genes encoding components of metabolic pathways in liver from ad libitum–fed mice. CR at night ameliorated these aging-related changes. Our results show that circadian interventions promote longevity and provide a perspective to further explore mechanisms of aging.

Caloric restriction (CR) without malnutrition or starvation, which is achieved by reducing ~30% of daily food intake, is the most effective nonpharmacological intervention that improves life span in model organisms (*1*), but the underlying mechanisms remain unclear (*2*–*6*). Classical CR protocols in mice lead to a temporal restriction of food intake with a long (>22 h) fasting interval because mice consume the food as soon it becomes available (*7*–*9*). Timed food administration is a potent signal that entrains circadian clocks in peripheral tissues such as liver (*10*–*12*). Thus, in addition to reducing daily energy intake, CR resets complex circadian programs of gene expression in tissues throughout the body (*13*–*15*). Although decreased energy intake is commonly thought to be the critical factor that extends life span, it is possible that the timing of food intake is also a key component. The changes caused by time-restricted feeding can have profound effects on physiology (*16*). For example, mice (which are nocturnal) fed a high-fat diet only during the day gained significantly more weight than mice fed the same diet only during the night (*17*). Also, mice fed a high-fat diet restricted to an 8-hour window during the night were protected against diet-induced obesity, hepatic steatosis, hyperinsulinemia, and inflammation compared with mice fed ad libitum (AL) (*18*, *19*). Thus, temporally restricted feeding at night, which is the normal active and feeding time of day for mice, is beneficial.

[1]Department of Neuroscience, Peter O'Donnell Jr. Brain Institute, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. [2]Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. [3]Animal Resources Center, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA.
*Corresponding author. Email: carla.green@utsouthwestern.edu (C.B.G.); joseph.takahashi@utsouthwestern.edu (J.S.T.)
†Present address: Infectious Diseases and Vaccinology, School of Public Health and Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA.

Although the timing of food intake can have an impact on health, it remains unclear whether the timing and frequency of feeding also affect life span in mice (*16*, *20*, *21*). Food consumption triggers behavioral and metabolic changes in mammals that have profound impacts on health status (*22*). We studied the contributions of feeding time and fasting under CR and compared behavioral, metabolic, and molecular outcomes throughout the life span. We tested five different CR protocols and an AL control group using automated feeders (*7*). After 6 weeks of baseline AL food access, C57BL/6J male mice were subjected to 30% CR. Mice were fed nine to ten 300-mg food pellets containing 9.72 to 10.8 kcal every 24 h starting at the beginning of the day (CR-day) or night (CR-night), similar to classical protocols in which mice consumed their food within 2 h as one meal (*7*). To prevent the 2-h binge-eating pattern and to reduce the fasting interval to ~12 h, two additional CR groups of mice were fed a single 300-mg pellet (1.08 kcal) delivered every 90 min to distribute the food access over a 12-h window either during the day (CR-day-12h) or during the night (CR-night-12h). A fifth CR group of mice was fed a single 300-mg pellet every 160 min continuously spread out over 24 h (CR-spread) to abolish the rhythmic pattern of food intake and to prevent any fasting intervals (Fig. 1A).

Behavioral and body weight dynamics with age

To select the diet for the longevity studies, we first compared standard laboratory chow (Teklad Global 2018) with two different precision food pellets with similar caloric content but different compositions: a grain-based diet that we used previously (F0170) (*7*) and a purified diet (F0075) (fig. S1A). We found that mice fed the purified diet showed body weight gain similar to the standard laboratory chow; however, the mice fed grain-based pellets gained significantly more weight (fig. S1B). Because the composition of grain-based diets is known to vary by batch and by season of the year (*23*) and because longevity experiments require at least 4 years, we chose the purified diet that could be completely defined and maintained over the entire duration of the life-span experiments and did not cause excessive weight gain compared with standard laboratory chow. We used an automated feeding system (*7*) and monitored feeding and wheel-running activity of individually housed mice continuously throughout their life span. This allowed us to measure behavioral and metabolic changes in mice under all six feeding conditions as they aged.

In agreement with previous studies (*24*), mice under unrestricted feeding (AL) gradually increased their body weight until 20 months of age, after which they showed an age-related decline (Fig. 1B and fig. S2). All CR groups maintained lower body weights throughout their life span, consistent with lower food intake (Fig. 1B and fig. S3). We previously showed that CR-day mice gained more weight than CR-night mice with the grain-based diet (*7*), but this effect was not reproduced with the purified diet (fig. S2), perhaps because of the difference in fat source in the two diets (fig. S1A). Long-term recordings of feeding events showed that mice adjusted their feeding patterns to match the externally controlled availability of food (including daytime feeding and 24-h spread-out feeding). These feeding patterns were consistently maintained throughout their life span (Fig. 1, C and E; fig. S4A; and data S1). Mice in the AL group normally consumed ~75% of their food at night and maintained this pattern of food consumption throughout their life span, with a gradual increase in food consumption with age after the first year (Fig. 1F and fig. S4A). Mice in the CR-night-2h and CR-day-2h groups with 24-h access to 30% CR (relative to AL controls for the first 200 days of study) rapidly consumed their daily allotment within 2 h, as previously described (Fig. 1, C and E) (*7*), and this 2-hour intake pattern was maintained throughout their life span (fig. S4A). Although AL mice increased their food consumption after 1 year of age, the amount of food was not increased for the CR groups (fig. S3), so the CR increased from 30 to ~40% compared with AL at later ages. Similarly, animals exposed to CR with food access spread over 12 or 24 h also adapted to the imposed meal pattern by eating each pellet as soon as it became available, which was every 90 min (CR-day-12h and CR-night-12h) or every 160 min (CR-spread) (Fig. 1, C and E, and figs. S3 and S4A). When examining the median phase of feeding, which is the time at which mice ate 50% of their daily allotment, we observed that the phase for classic CRs

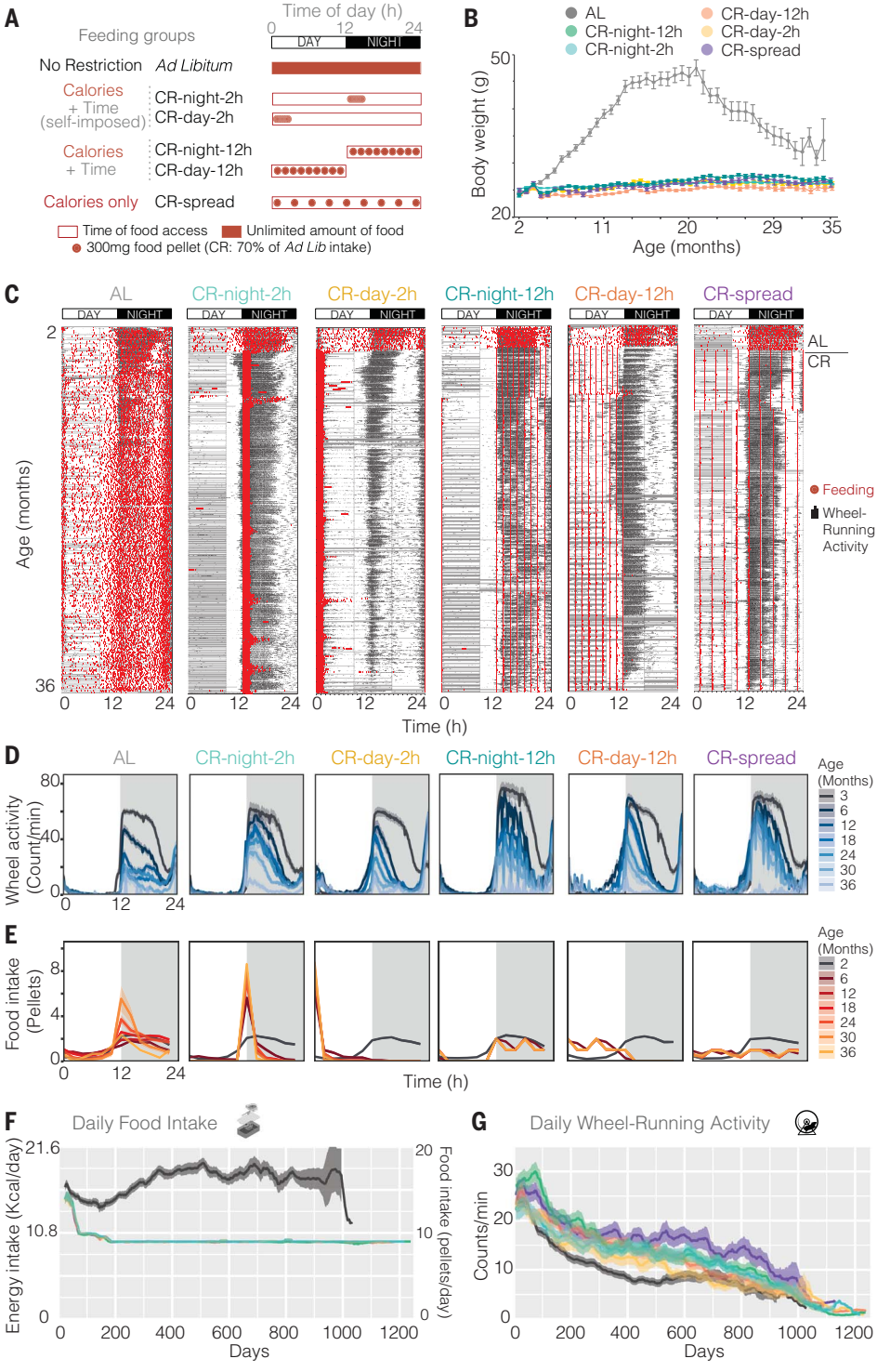
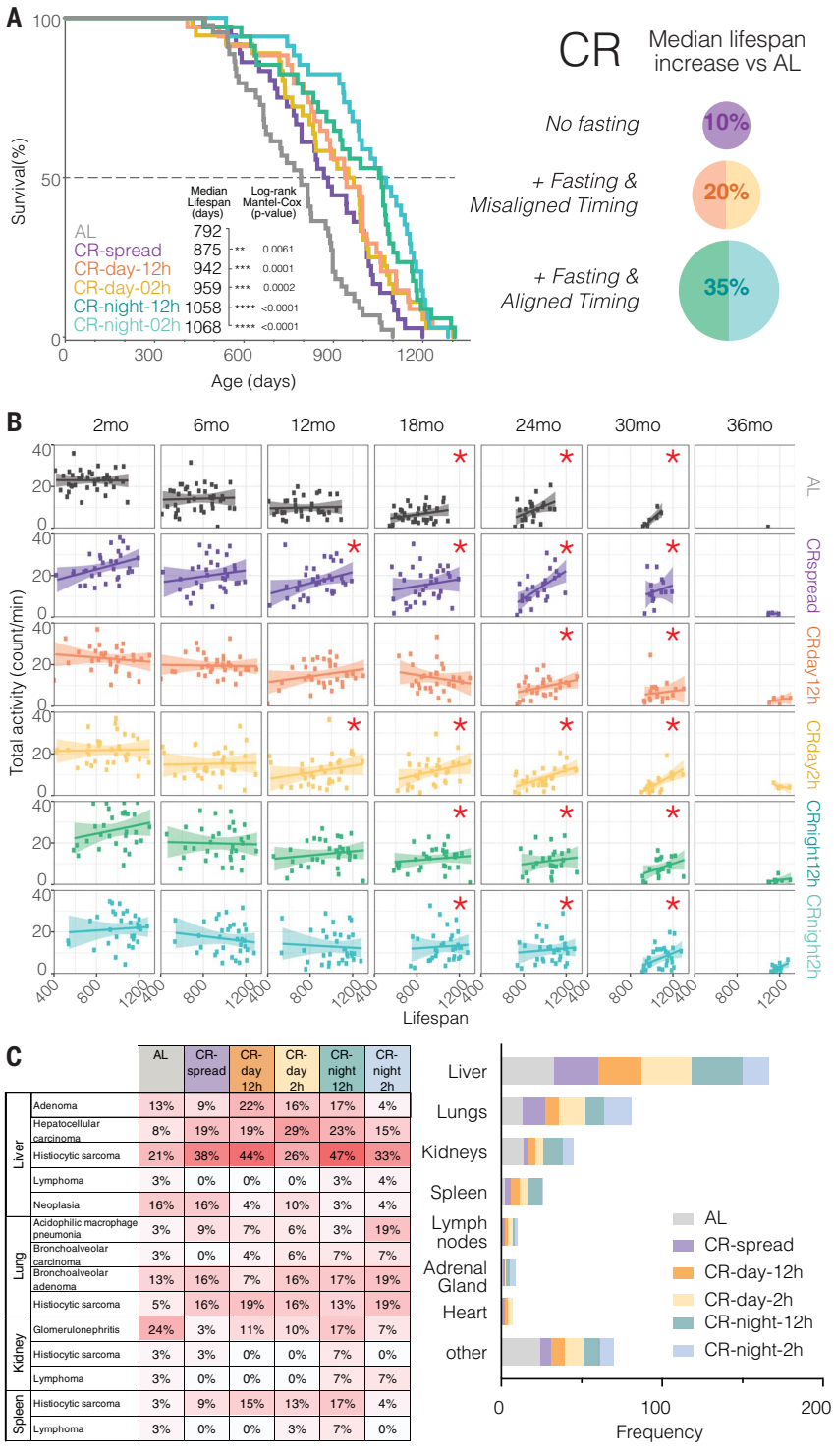


Fig. 1. Effects of CR on body weight, circadian behavior, and feeding in C57BL/6J male mice. (**A**) Experimental design showing feeding conditions for each of the six groups. (**B**) Average body weights (±SE) of mice in all six groups (n = 43 for the AL group and n = 36 for each of the CR groups) taken every 3 weeks throughout the experiment. (**C**) Examples of double-plotted actograms from each experimental group overlaying wheel-running (black histograms) and feeding (red dots) behaviors. All mice were on AL feeding for the first 6 weeks (period above the line on the right of the actograms) before the CR began. (**D** and **E**) Twenty-four-hour profile of wheel-running activity (D) and food intake (E) at different ages (averaged over 21 days, n = 36 to 43 mice) for each group. (**F**) Energy intake per day (left axis) and number of food pellets per day (right axis) for each group throughout the experiment. For the AL group, the dark line is average and the gray shading is SE. All CR groups were limited to 70% of AL consumption for the first 200 days of age and were not adjusted after 200 days, so no variation is observed. (**G**) Daily wheel-running activity (average counts/min over 24 hours ± SE) throughout the experiment.

was 1 h after the food onset [Zeitgeiber time (ZT) 1 h for CR-day-2h and ZT 13 h for the CR-night-2h]. For the CR-spread group, the phase was ZT12 because the mice ate equal amounts during the day and night (fig. S4A). The feeding pattern was also consistent with daily changes in body weight (fig. S4B). With the exception of the CR-spread group, in which the food was equally distributed throughout the day, body weight changed throughout 24 h with a significant increase during the feeding time (fig. S4B). This finding was more pronounced in classic CR protocols, with the highest body weight gain of 3 g occurring between ZT0 and ZT4 in CR-day-2h mice and between ZT12 and ZT16 in CR-night-2h mice. These body weight gains are consistent with the observation that mice eat their entire allotment (2.7 to 3 g) as one single meal within 2 h.

All groups maintained a normal nocturnal locomotor activity pattern for life, with the exception of the day-fed mice, which tended to have more daytime activity (Fig. 1, C and D). Overall, these long-term recordings showed that when food was restricted to the daytime, mice interrupted their "rest phase" to eat but



		AL	CR-spread	CR-day 12h	CR-day 2h	CR-night 12h	CR-night 2h
Liver	Adenoma	13%	9%	22%	16%	17%	4%
	Hepatocellular carcinoma	8%	19%	19%	29%	23%	15%
	Histiocytic sarcoma	21%	38%	44%	26%	47%	33%
	Lymphoma	3%	0%	0%	0%	3%	4%
	Neoplasia	16%	16%	4%	10%	3%	4%
Lung	Acidophilic macrophage pneumonia	3%	9%	7%	6%	3%	19%
	Bronchoalveolar carcinoma	3%	0%	4%	6%	7%	7%
	Bronchoalveolar adenoma	13%	16%	7%	16%	17%	19%
	Histiocytic sarcoma	5%	16%	19%	16%	13%	19%
Kidney	Glomerulonephritis	24%	3%	11%	10%	17%	7%
	Histiocytic sarcoma	3%	3%	0%	0%	7%	0%
	Lymphoma	3%	0%	0%	0%	7%	7%
Spleen	Histiocytic sarcoma	3%	9%	15%	13%	17%	4%
	Lymphoma	3%	0%	0%	3%	7%	0%

Fig. 2. Extent of CR-mediated increases in longevity depend on feeding time. (**A**) Survival curves for each group (n = 43 for AL and n = 36 for each of the CR groups) are shown in the left panel and the median life span (days) is shown in the inset. Right panel summarizes the results, showing the increase in life span from timed feeding with the largest increase when food is restricted to night. (**B**) Correlation plots comparing life span (days) for each mouse with its daily averaged total activity (counts/min) at different ages. Increased activity significantly correlates with longer life span in old, but not young, mice in all groups (see asterisks, Spearman correlation). (**C**) Necropsy followed by histopathology results showing pathologies and diseases (left) and tissues mostly affected (right) at the time of death for each feeding condition.

maintained most activity during the nighttime (Fig. 1, C and D). Therefore, feeding and locomotor activity are misaligned with daytime feeding in these animals throughout their life span, which would be expected to lead to adverse metabolic consequences (*25*, *26*). Activity of the mice declined as they aged (*27*), with AL mice having the lowest activity levels compared with the CR groups between 6 and 18 months of age [Fig. 1G and fig. S5; two-way analysis of variance (ANOVA); age, $P < 0.0001$; feeding, $P < 0.0001$; interaction nonsignificant (NS)].

Life-span extension by CR depends on feeding time

We investigated the contribution of calories, feeding time, and fasting period on longevity. CR was sufficient to extend the median life span in male mice, but the range of this extension depended on when the food was consumed (Fig. 2A). The percentage of life-span extension varied across conditions. Consistent with other reports, AL mice had a median life span of 792 days (*24*, *28*). CR-fed mice lived 10 to 35% longer than AL mice depending on the CR group. The CR-spread group, which had a 30% reduction in calories but with feeding spread throughout the day-night cycle, had a median life span of 875 days, which is 10.5% longer than that of AL mice, demonstrating that CR alone without time restriction or fasting is sufficient to extend longevity. The CR-day-12h and CR-day-2h groups had median life spans of 942 and 959 days, respectively, which are 18.9 and 21.1% longer than the life spans of AL mice, respectively. Thus, in addition to the reduction in calories, a minimum of 12 hours of fasting induces its own benefits on longevity. There were no significant differences in life span when day-fed mice fasted for ~22 hours versus 12 hours, indicating that 12 hours of fasting is sufficient. CR-night fed mice outlived both CR-day groups: The CR-night-12h and CR-night-2h mice had median life spans of 1058 and 1068 days, respectively, which are 33.6 and 34.8% longer than the life spans of AL mice, respectively. Again, there was no additional benefit of ~22 hours of fasting compared with 12 hours of fasting in these groups fed at night, indicating that 12 hours of fasting is sufficient for prolonging life span. There was a significant extension of life span by CR-night-2h mice (34.8% extension) over CR-day-2h mice (21.1% extension) (log-rank Mantel-Cox, $P < 0.05$), which differed in the relative phase of food consumption by the mice. It is possible that sleep disruption due to misaligned feeding could contribute to the difference in life span. Further studies are required to determine whether sleep is affected and, if so, if potential sleep disruptions can contribute to the differences in life span observed between CR-day

versus CR-night animals. However, it has recently been shown that sleep homeostasis is maintained in mice under a restricted feeding schedule in which food is only available for 4 h in the middle of the "sleep" phase (ZT4 to ZT8) (*29*). In all five of the CR groups in this study, the mice consumed exactly the same number of daily calories throughout their life span (figs. S3 and S4; two-way ANOVA; age, $P < 0.0001$; feeding, $P < 0.0001$; interaction, $P < 0.0001$), yet the pattern and circadian phase of feeding had major effects on life span. A >12-hour fasting interval combined with nocturnal (normal) feeding yielded the greatest benefits on life span. Thus, calories are processed differently depending on when they are consumed, and anti-aging interventions such as CR can be optimized by timing them to a specific time of day. Maximum life span, estimated as the 10% longest-lived mice in each group, was significantly longer in all of the CR groups compared with AL except for the CR-spread group (exact Fisher's test, $P < 0.05$) (data S1). Among the CR groups, only CR-night had a significant increase in maximum life span compared with CR-spread (exact Fisher's test, $P = 0.0256$ CR-night versus CR-spread). This suggests that feeding and fasting cycles that are in sync with internal circadian clocks (CR-night-2h) extend both the median and maximum life span, which is indicative of delaying the aging process as opposed to delaying the onset of a single disease.

Necropsy followed by histopathology revealed that all groups had similar diseases at death, but in the CR groups, these diseases occurred at older ages (coinciding with longer life spans) as previously reported (*8*). Neoplasias were the most frequent pathology in all groups, with histiocytic sarcomas being the major cause of death, followed by hepatocellular carcinoma (Fig. 2C and data S1). Histopathology analysis of the target tissues also revealed the highest incidence of lesions in the liver (Fig. 2C).

Age-related decline in activity predicts lower survival in mice

To evaluate whether behavioral or metabolic parameters correlated with longer life spans, we compared feeding (total daily intake), body weight, and wheel-running activity (total daily activity and percentage nighttime activity) with life span. We found no correlation of life span with food intake or body weight at any age (figs. S6 and S7). However, in all feeding conditions, daily locomotor activity level positively correlated with longer life span after 18 months of age (Fig. 2B). Additionally, higher activity levels during the normal circadian phase (at night) after 24 months of age also positively correlated with longer life span (figs. S8 and S9). Because voluntary wheel-running activity does not affect life span in mice (*30*), our results suggest that the level of wheel-running activity after 18 months of age could be a biomarker for health span. Thus, the activity level of mice at older ages (>18 months) can be used as a predictor of longer life span.

CR promotes widespread metabolic benefits

Body composition analysis at 12 and 20 months of age showed that although fat mass was significantly higher in the AL group versus the CR groups, there were no differences in fat mass among the CR groups (fig. S10). We assessed metabolic markers from plasma at 6 and 19 months of age. Insulin levels increased with age under AL, and such increases were attenuated by all CR groups, which maintained low insulin levels at both ages (fig. S11). In young mice, CR groups had similar insulin levels yet lower glucose levels in plasma compared with the AL group (figs. S10 and S11), suggesting that improved insulin sensitivity was associated with lower food intake. As the mice aged, similar levels of circulating glucose were found in all feeding groups even though AL-fed mice had higher insulin levels than CR groups, suggesting that CR generally protects against age-related insulin resistance. Similar to what was seen with insulin, there was a significant age-related increase in leptin under AL that did not occur in the CR groups. Although all CR groups maintained lower leptin levels than the AL group at 19 months of age, only the CR-night groups had leptin levels that were significantly lower at younger ages, suggesting that alignment of feeding may play a role in regulating satiety at both ages. Glucagon-like peptide 1 (GLP-1) is an intestinal hormone that is secreted after meals and decreases blood sugar levels by promoting insulin secretion and suppressing glucagon release (*31*). We found that although the levels of GLP-1 did not change with aging, the longest-lived (CR-night) groups had significantly lower levels at older ages compared with AL (fig. S11). This indicates that CR-night mice may have had an improved sensitivity to GLP-1 in regulating insulin secretion and glucose levels. Overall, the longest-lived CR groups had improved hormonal profiles, insulin sensitivity, and glucose homeostasis as they aged.

Gene expression changes with aging and CR

To determine the effects of aging and feeding at the molecular level, we assessed circadian gene expression patterns using RNA sequencing (RNA-seq) in mouse liver in all six feeding conditions at two ages: young mice at 6 months and old mice at 19 months. We chose to profile the liver because it is a major metabolic target of the circadian clock system (*32*) and because it had the highest incidence of age-related lesions among target tissues (Fig. 2C). We chose 6 months of age to assess fully mature adult mice and 19 months of age to assess aged mice before their obvious decline in body weight (Fig. 1B) and before there was >10% mortality in AL mice (Fig. 2A). In each group, we profiled the liver at 12 time points every 4 hours for 48 hours across two circadian cycles in mice transferred to constant darkness to assess circadian rather than diurnal cycles (two biological replicates × 12 time points = 24 liver samples per feeding condition). We chose to perform these gene expression experiments in constant darkness to determine whether a rhythmic gene profile is circadian. Because liver-cycling gene expression patterns are strongly driven by feeding cycles (*33*), the absence of light:dark (LD) cycles would be expected to yield results for rhythmic gene expression in the liver similar to those seen in LD cycles reported previously (*11*, *33*). Unbiased principal component analysis showed that (i) samples from young and old groups clustered separately under AL, (ii) young groups under CR clustered together and separately from AL, and (iii) old groups under CR clustered together between aged-AL and young-CR groups (Fig. 3A). This suggests that at the molecular level, the aging process is different between mice fed AL or CR, and aged-CR mice maintained a liver gene expression pattern more similar to that of the young animals.

To address the overall impact of aging at the molecular level, we performed differential gene expression analysis between young and old mice (data S2 and S3). A total of 2599 genes (18.6% of expressed genes) were differentially expressed with aging under AL feeding (Fig. 3B and fig. S12). Of these, 2031 genes were up-regulated and 568 were down-regulated. Gene ontology (GO) analysis revealed that the up-regulated genes were highly significantly related to immune system processes and inflammation (Fig. 3C), whereas the down-regulated genes were related to metabolism (Fig. 3D and data S4). This is consistent with previous reports showing that increased inflammation and senescence are hallmarks of aging (*3*) and recent work on glycine-serine-threonine metabolism in longevity (*34*). Among the up-regulated genes were *Cd36* ($\log_2$FC = 3.8, adjusted $P = 2.95 \times 10^{-92}$), which is a multifunctional glycoprotein that acts as a receptor for ligands such as thrombospondin, fibronectin, amyloid beta, oxidized low-density lipoprotein, and long-chain fatty acids, among others (*35*), and the peroxisome proliferator-activated receptor gamma (*Pparg*, $\log_2$FC = 1.8, adjusted $P = 1.5 \times 10^{-28}$), a transcription factor also associated with immunosenescence during aging (Fig. 3C) (*36*, *37*). Other genes of note were *Adora1*, *Serpine1*, *Themis*, 10 Toll-like receptor (*Tlr*) genes, *Spon2*, and *Zcchc11* (data S2 and S3). Among the down-regulated genes, there was an enrichment in key enzymes of amino acid (*Gnmt* and *Agxt*) and cholesterol metabolism such as HMG-CoA reductase

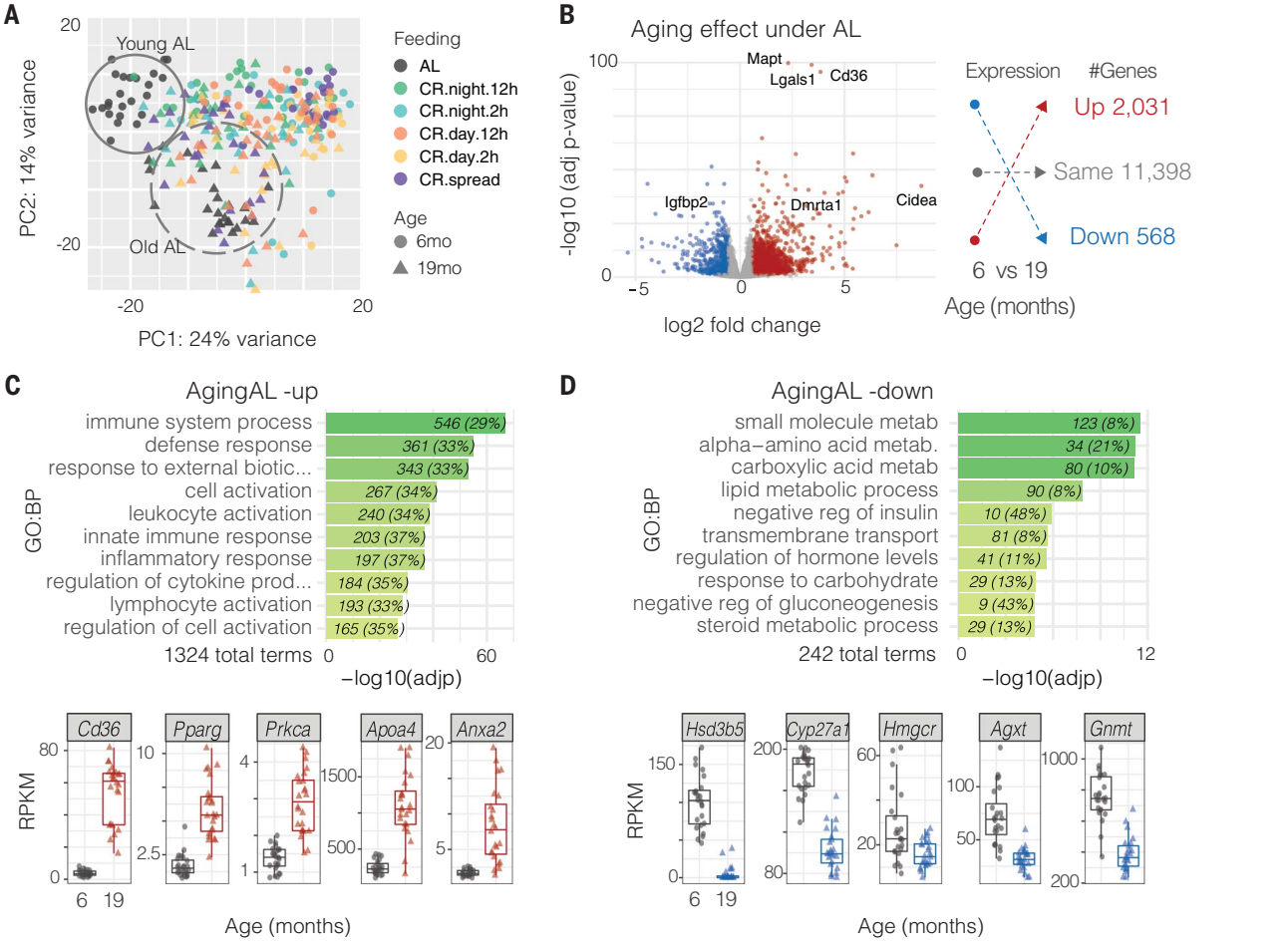


Fig. 3. Gene expression signatures in liver change during aging in AL mice. (**A**) Principal component analysis of gene expression from liver mRNA-seq. mRNA-seq data are from 48 mice for each feeding condition (n = 24 from 6 months of age, n = 24 from 19 months of age), with livers collected every 4 hours over 48 hours while mice were in constant dark. Circles indicating young AL mice (solid line) are clustered together, and triangles indicating aged AL mice (dashed line) are in a distinct cluster. Liver gene expression data among CR groups cluster together independently of age. (**B**) Volcano plot showing differential gene expression in young versus aged AL mice. Red denotes genes for which expression is significantly increased in aged AL mice; blue denotes genes that are significantly decreased in old AL mice. The number of mRNAs in each category are shown in the right panel. (**C**) GO terms of genes that are increased in aged AL mice (top) and examples of gene expression (bottom). Gray indicates young AL, and red indicates aged AL. (**D**) GO terms of genes that are decreased in aged AL mice (top) and examples of gene expression (bottom). Gray indicates young AL, and blue indicates aged AL.

(*Hmgcr*), which is implicated in liver disease and hepatocellular carcinoma (*38*–*41*) (Fig. 3D). Other metabolic genes of note were *Got*, *Lepr*, *Lpin1*, *Pfkfb3*, *Scap*, *Hsd17b2*, *Hsd3b5*, as well as 14 cytochrome P450 (*Cyp*) genes and 28 solute carrier (*Slc*) genes (data S2 and S3). Overall, there was an increase in the expression of inflammatory genes and a decrease in the expression of metabolic genes.

CR alone rescues most age-related changes observed under AL

To determine the overall effect of CR, we analyzed which genes underwent age-related changes in any condition and identified those with expression levels that remained protected under CR versus AL. Across all feeding conditions, there were 4077 genes with expression that was up- or down-regulated with aging (Fig. 4A and data S2), indicating that 29% of the liver transcriptome is susceptible to aging-related changes under any condition tested. AL-fed mice had the highest percentage of genes that changed with aging (18%), whereas all of the CR groups had lower overall changes in gene expression with age (Fig. 4A). The CR-night-2h group had the smallest overall change in gene expression with age (4%) (Fig. 4A). Age-related fold change of individual genes in the CR-spread and CR-day groups were partially decreased compared with the fold changes seen in AL, whereas CR-night-2h strongly attenuated these age-related changes. Figure 4B shows these results as correlation plots of age-related fold change compared with AL for all CR groups. If there were no rescue by CR, the data points would fall on the unity line (slope = 1), whereas with complete rescue by CR, the data points would fall on the horizontal line (slope = 0). As seen in Fig. 4B, the regression line for the CR-night-2h group was almost completely flat, demonstrating that CR-night-2h strongly reduced age-dependent changes in gene expression. Approximately 50% of age-related changes in gene expression under AL (1233 genes) were restored in every CR condition (Fig. 4C and data S2) by rescuing 44% of up-regulated genes (inflammation and immune function) and 60% of down-regulated genes (metabolic pathways). This result indicates that CR alone prevents most of the age-related changes observed in the control AL condition.

GO analysis showed that the genes protected by CR were also associated with immune function, inflammation, and metabolism (Fig. 4C and data S4) (*34*). Among these genes were those

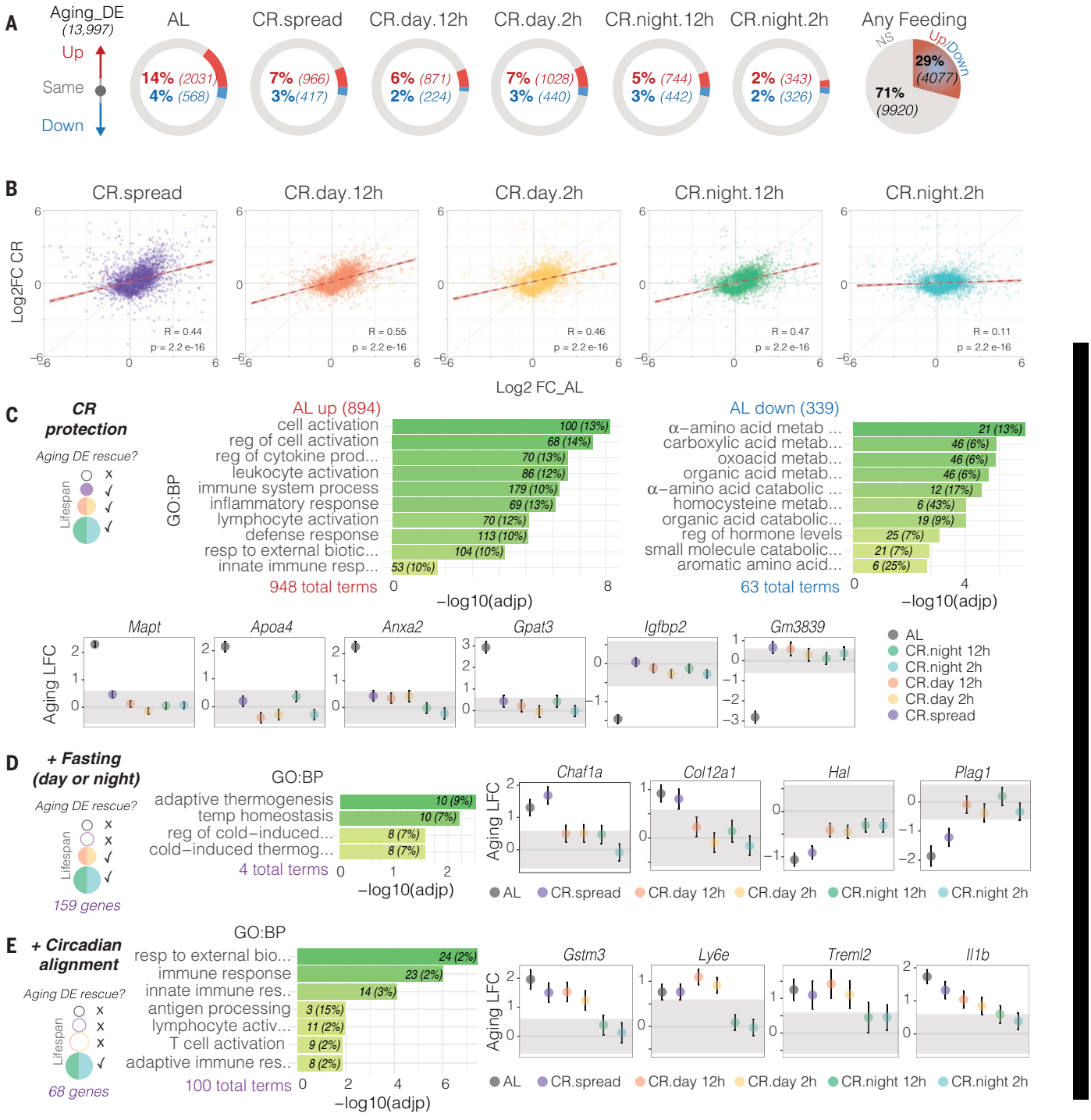


Fig. 4. CR ameliorates age-related changes in liver gene expression observed under AL conditions. (**A**) Schematic comparison of differential gene expression between young and old mice in the six feeding conditions. Circles show the percentage of genes that are unchanged between young and aged mice (gray), increased in aged mice (red), and decreased in aged mice (blue). Pie chart on the right shows the percentage of genes susceptible for age-related changes in any feeding condition. (**B**) Spearman correlation plots comparing changes in gene expression between the aging DE genes between AL and CR groups (aging DE genes are defined here as the 4077 genes that change with age in any of the six feeding conditions tested). (**C**) Schematic representation of genes that are protected from age-related changes in every CR group (left) with GO terms of those significantly up-regulated (middle) or down-regulated (right) in aged AL mice. Represented are 10 nonredundant of the top 25 most significant enriched terms. Examples of age-related fold changes ($log_2FC \pm SE$) in gene expression are shown below for all feeding conditions. Gray-shaded areas indicate FC < 1.5 considered as not significant change. (**D** and **E**) Schematic representation, GO, and representative genes that maintain similar levels between young and old ages due to fasting (day or night) (D) and circadian alignment of feeding and fasting cycles (E).

encoding for microtubule-associated protein tau (*Mapt*) and apolipoprotein A-IV (*Apoa4*), the expression of which was up-regulated in old-AL mice but maintained at lower (young) expression levels in all CR groups (Fig. 4C). These genes have been linked with aging and neurodegeneration in Alzheimer's disease (*3*, *42*). A similar aged-related up-regulation occurred in the liver and was reduced by CR. Of the metabolic genes that declined with age under AL, such as insulin-like growth factor-binding protein 2 (*Igfbp2*), CR strongly attenuated changes in expression in all five CR groups.

Fasting-related genes

To evaluate gene expression changes caused by fasting, we compared genes that were differentially expressed only in AL and CR-spread but remained constant in all the other four CR groups with some degree of fasting. We found 159 genes of this type that could be potential candidates accounting for the change from 10 to 20% life-span extension (Fig. 4D and data S2). Among these were collagen type XII alpha 1 chain (*Col12a1*), which is associated with cancer (*43*) and is elevated in liver fibrosis, and pleomorphic adenoma gene 1 (*Plag1*), an oncogene associated with hepatoblastoma and age-related decrease in skeletal muscle (*44*, *45*). Other genes included chromatin assembly factor 1 subunit A (*Chaf1a*) (*46*) and histidine ammonia-lyase (*Hal*) involved in amino acid metabolism. GO analysis revealed that these fasting-related genes were also enriched for thermogenesis pathways.

Time-related genes

To evaluate the beneficial effect of feeding time, we searched for genes that maintained similar levels at young and old ages only in the CR-night fed groups with the longest life spans but were differentially expressed in the AL, CR-spread, and CR-day fed groups. We found 68 genes that were specifically protected in the CR-night groups (Fig. 4E and data S2). GO analysis revealed specific subsets of genes involved in the immune system and inflammation, such as glutathione S-transferase mu 3 (*Gstm3*), which protects against oxidative stress; lymphocyte antigen 6 family member E (*Ly6e*), which regulates T-cell proliferation, differentiation, and activation; triggering receptor expressed on myeloid cells-like 2 (*Treml2*); and proinflammatory cytokines such as interleukin-1β (*Il1b*) that are increased in the elderly (*3*). Thus, circadian alignment of feeding time adds another level of protection of immune function and age-related inflammation beyond that seen with CR and fasting.

Circadian cycling of gene expression with aging

Robust circadian rhythms are at the core of a healthy physiology, and rhythm amplitude decreases in response to infectious diseases and aging (*22*, *47*, *48*). We investigated how feeding conditions, timing, and CR influenced circadian cycling of gene expression in young and aged mice. To search for circadian cycling genes, we performed RNA-seq from liver samples collected every 4 hours for 48 hours (data S4 and S5). We used very strict criteria to identify circadian cycling genes by selecting only those genes that were significant in three out of three different algorithms (JTK_CYCLE, ARSER, and RAIN) with a P value < 0.05 and a false discovery rate (FDR) < 5% (FDR < 0.05). We found 1718 rhythmic genes in young AL mice and 1507 rhythmic genes in old AL mice (see Fig. 5A, heatmaps of cycling genes, and data S5). The overlap of these two genes sets was 694 genes. Therefore, circadian cycling genes were both lost and gained with age (Fig. 5A). Figure 5B illustrates six cycling genes in young and old mice. Of the four circadian clock genes, *Arntl*, *Nr1d1*, *Per1*, and *Per2*, three had a lower amplitude in old mice. Two metabolic pathway genes, *Gys2* and *Pck1*, also had lower amplitude with age consistent with the overall decline in average gene expression in metabolic pathways (Fig. 3, B and D; see fig. S13 for example circadian profiles of pro-aging and pro-longevity genes). We compared the circadian phase and amplitude (fold change between trough and peak on the first and second circadian cycle) for the 694 shared cycling genes in young and old mice (Fig. 5A). There was no change in the phases of cycling genes with age; however, the amplitude of these cycling genes was lower, as seen by the deviation of the regression line from unity (slope = 0.5941 ± 0.009, P < 0.0001) (Fig. 5C). GO analysis of these cycling gene sets showed enrichment for metabolism, cell communication, signaling, and circadian rhythms (Fig. 5, D and E, and data S3), which is consistent with the decline in average gene expression in metabolic pathways (Fig. 3D) and with analysis of circadian gene targets using chromatin immunoprecipitation sequencing, in which the GO category metabolism was highly significant (*32*).

Effects of day versus night CR on circadian gene expression

Because the median life span of mice in the CR-night-2h and CR-day-2h groups was significantly different (1068 versus 959 days, respectively, and log-rank Mantel-Cox, P < 0.05), we focused on differential gene expression in these two CR groups. One essential difference in these two groups is the phase of food consumption relative to the LD cycle and relative to the circadian phase of the mice as assessed by their locomotor activity rhythms. We selected the sum total genes that showed circadian cycling in at least one of these four groups: CR-day-2h and CR-night-2h from young and old mice, which led to a total of 1491 cycling genes. Figure 6A shows cycling genes in these four groups as a heatmap in which each line represents the color-coded levels of expression of one gene (in each row) across time points (columns). There was a slight reduction in cycling genes in young CR-day-2h mice, but in old CR-day-2h mice, there were only seven cycling genes. Circadian gene expression profiles of the six genes shown in Fig. 5C showed that the day-fed groups at 19 months of age have either opposite phases (*Arntl*, *Nr1d1*, *Gys2*, *Per2*, and *Pck1*) or disrupted circadian profiles (*Per1*) (see Fig. 6B and fig. S13 for example circadian profiles of pro-aging and pro-longevity genes). We compared the fold-change amplitude of the 1491 genes in young and old mice. At both 6 and 19 months of age, the CR-night-2h groups had a significantly higher amplitude compared with the CR-day-2h groups, as seen in the fold-change frequency histograms and correlation plots (Fig. 6C). Thus, CR-night-2h feeding enhanced circadian amplitude relative to that of CR-day-2h.

To assess the phase of entrainment in these four CR groups, we compared the phases of cycling genes relative to those of AL mice at 6 and 19 months of age. Figure 6D shows correlation plots of genes that overlapped with AL mice in either the CR-day-2h or CR-night-2h groups at 6 and 19 months of age. The CR-night-2h mice shared many more cycling genes with AL mice than did the CR-day-2h mice at both ages. In addition, the phases of gene expression of CR-night-2h mice of both ages were correlated with the phases of the genes from the respective AL age group, with the data points falling on the unity line. As expected, the CR-day-2h mice had opposite or divergent phases relative to those of AL mice. For the old CR-day-2h mice, only seven genes were scored as cycling and only five overlapped with those of AL mice. Thus, in CR-day-2h mice, there was a paucity of cycling genes and this declined precipitously with age. A similar reduction was seen in fold-change gene expression in CR-spread mice compared with CR-night-2h mice at 6 and 19 months of age (fig. S14). The CR-night-2h groups showed robust circadian cycling, even in the old mice, suggesting that this intervention is very effective in rescuing circadian cycling of gene expression relative to CR-day-2h. A proviso to this cycling analysis is that these cycling algorithms, JTK_CYCLE and ARSER, are biased toward sinusoidal waveforms, and although RAIN can search for spiky or sawtooth waveforms, our requirement that a cycling gene is significant (FDR < 0.05) in all three algorithms results in only genes with sinusoidal waveforms passing this threshold. Daytime feeding and 2-hour feeding bouts both modified the gene expression time series waveforms so that they were less sinusoidal, and this contributed to the lower number of genes

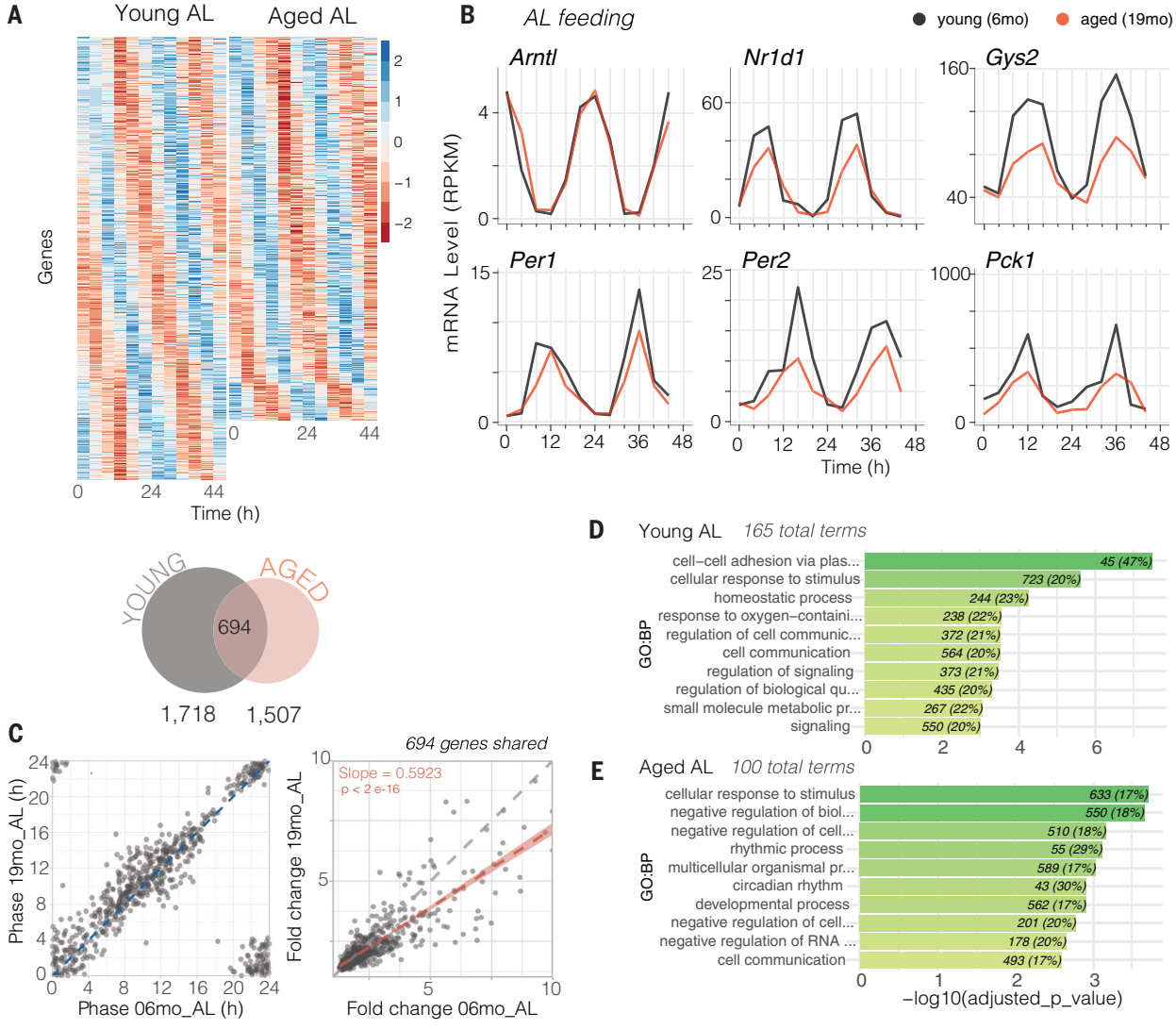


Fig. 5. Circadian rhythms in liver gene expression are blunted during aging in AL mice. (**A**) Gene expression patterns from mRNA-seq were analyzed for circadian rhythms using the ARSER, JTK_CYCLE (from Metacycle R Package), and RAIN circadian algorithms. Heatmaps (top) are sorted by phase of gene expression. Each row is one gene with expression level in *z*-score at 12 time points (columns). Venn diagram (bottom) shows the number of rhythmic genes in young (gray) and aged (red) AL livers using stringent criteria (significantly cycling according to three algorithms; Benjamini-Hochberg *P* and q < 0.05 and $\log_2$FC > 0.3) to define rhythmicity. (**B**) Examples of circadian profiles of genes that are rhythmic in both young and aged AL livers. Black indicates young AL livers, and red indicates aged AL livers. (**C**) Comparison of phase (left, hours) and amplitude (right, daily fold change) of the 694 genes that were rhythmic in both age groups. The red correlation line (Spearman) and linear regression (slope is statistically different from 1; $P < 2 \times 10^{-16}$) in the fold change comparison indicates that aged animals showed overall reduced amplitude of rhythmic genes. (**D** and **E**) GO terms of genes that are cycling in either young (D) or aged (E) AL mice. Represented are 10 nonredundant terms of the top 25 most significant enriched terms.

identified as cycling in the CR-day-2h group. Therefore, we do not recommend putting too much weight on the number of cycling genes called but rather emphasize the phase and amplitude of gene expression.

Discussion

Classic CR protocols not only reduce energy intake but also lead to severe behavioral time-restricted feeding behavior and prolonged fasting intervals (*7*–*9*). Because time-restricted feeding and fasting are beneficial to health, these two factors may contribute to life-span extension in classic CR experiments. In the work presented here, we deconvolved the effects of calories, fasting, and circadian alignment on longevity. We compared five different CR feeding groups that differed only in the daily pattern of food consumption without any changes in food composition or energy content. We used an automated feeding system (*7*) to test whether the timing and fasting period between meals affected life span in male mice under 30% CR by allowing food access only during the day, at night, or evenly distributed throughout 24 hours. By spreading out food through 24 hours, with no day-night feeding pattern in the CR-spread group, we found only an ~10% extension of median life span compared with AL-fed mice. The CR-day fed groups that had an ~12- or ~22-hour fasting interval lived ~20% longer than AL control mice. The degree of life-span extension was significantly longer when food was consumed during the nighttime, which is the normal feeding time in nocturnal rodents (~35% versus 20% compared with AL, log-rank Mantel-Cox $P < 0.0001$, and ~10% night versus day, log-rank Mantel-Cox $P < 0.05$) (Fig. 2A). Although potential sleep disruption needs to be carefully studied in CR-day versus CR-night groups, recent evidence shows that sleep homeostasis is

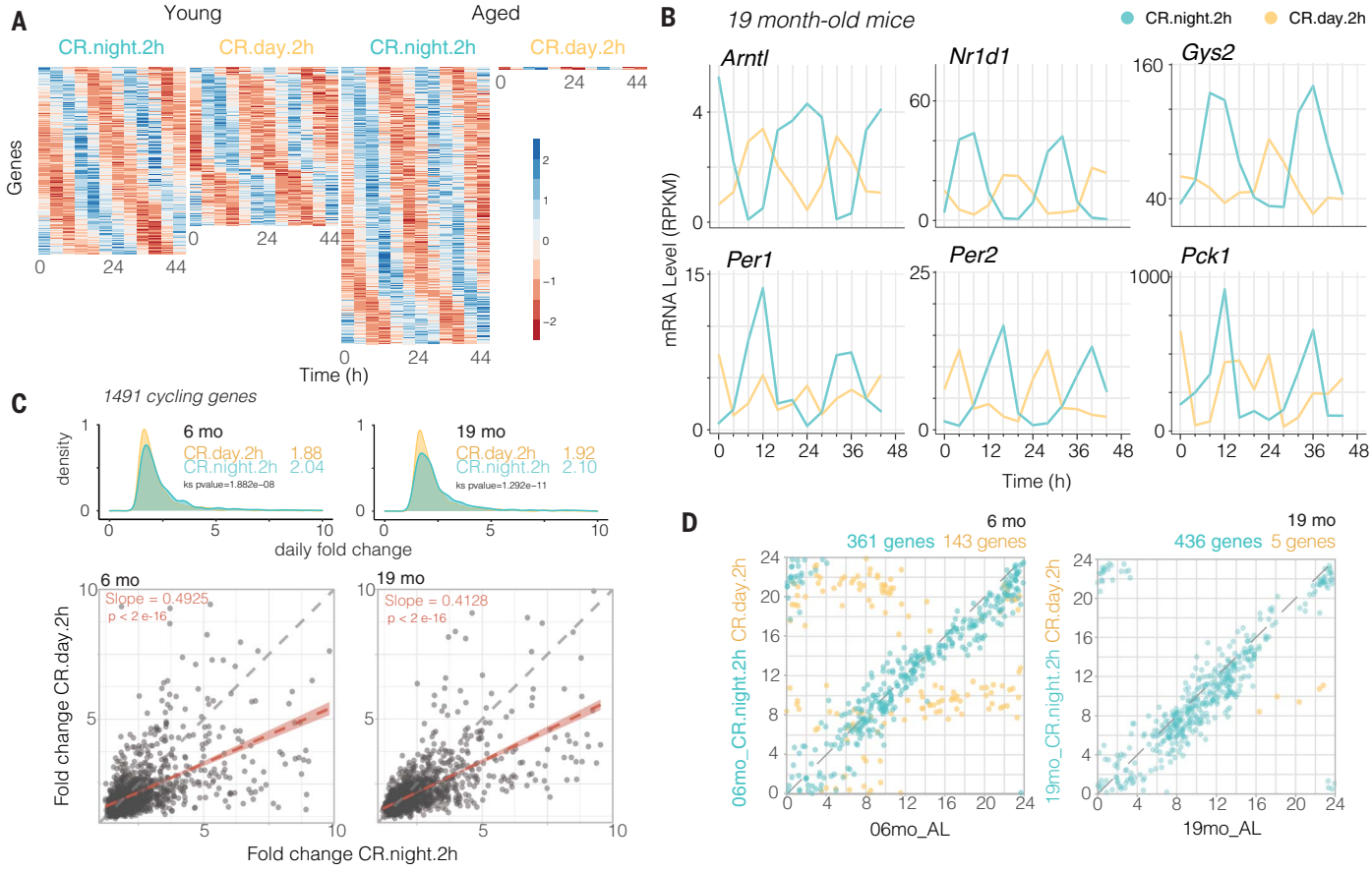


Fig. 6. Effects of CR and phase of feeding on circadian gene expression. (**A**) Gene expression patterns from mRNA-seq were analyzed for circadian rhythms using the ARSER, JTK_CYCLE (from the Metacycle R Package), and RAIN circadian algorithms. Heatmaps sorted by phase of gene expression. Each row is one gene with expression level in *z*-score shown at 12 time points (columns). (**B**) Examples of the circadian profiles of the same genes shown in Fig. 5B comparing profiles from CR-night-2h (blue) to CR-day-2h (yellow) aged mice. (**C**) Comparison of circadian amplitude (daily fold-change) of 1491 rhythmic genes from young (left) and aged (right) CR groups. Top panels show amplitude density plots (median amplitude values are inset). Bottom panels are correlation plots comparing amplitude of genes from CR-night-2h to CR-day-2h in young (left) and aged (right) mice. The linear regression lines have slopes that are significantly <1 ($P < 2 \times 10^{-16}$). (**D**) Phase correlation plots of rhythmic genes from young (left) and old (right) CR-night-2h-fed (blue) and CR-day-2h-fed (yellow) mice versus AL-fed mice of the same ages. Phase is represented in hours. Numbers of shared cycling genes between each CR condition and AL are labeled on top of the correlation plot.

maintained with daytime feeding (*29*). All CR groups maintained a relatively steady body weight throughout their life span, indicating that the additive effects of CR combined with appropriate timing of food consumption were independent of weight gain. Thus, the maximal pro-longevity benefits of CR can be achieved by a fasting interval of >12 hours in which a time-restricted feeding interval occurs in phase with the natural nocturnal circadian phase of feeding in mice (i.e., circadian alignment).

These results are consistent with recent studies in C57BL/6J male mice in which similar life-span extensions were reported in once-per-day CR mice fed in the morning (20%, equivalent to our CR-day-2h group) (*9*) or mice on ~28% CR when food was consumed during the daytime but closer to the normal active phase at night (9 hours later than our CR-2h-day and 3 hours earlier than our CR-2h-night groups) (*8*). Pak *et al.* (*9*) suggested that fasting alone drives the geroprotective effects of CR, which is partially consistent with our results; however, they argued that using a 50% cellulose-diluted diet that 30% CR does not extend life span and did not test whether the phase or circadian alignment of CR could be a factor. By contrast, we showed that 30% CR without dilution of the diet in the CR-spread group extended life span by ~10%; thus, we conclude that 30% CR alone without fasting or circadian alignment accounts for a 10% extension of life span. In our study, the diets used were identical in all six groups, which mitigates against the confounding effects of diet composition such as fiber (cellulose) and uncontrolled grain-based diets (*23*, *49*, *50*). Therefore, our conclusions differ from Pak *et al.* (*9*) in two ways: (i) we found that CR alone without fasting can still extend life span by ~10% using the same diet and (ii) we found that fasting contributes in an additive manner to life-span extension rather driving the effects of CR.

In nonhuman primates, the effects of CR on longevity have differed between studies performed at the University of Wisconsin (UW) and the National Institute of Aging (NIA). However, many differences in study design, diet, and feeding protocols have been documented, and the overall conclusions from a joint consensus is that CR improves health and survival in rhesus monkeys (*51*–*53*). CR did not extend life span in the NIA study; however, the AL control group was longer lived, slightly caloric restricted, food was provided twice a day, and body weights were lower than AL controls in the UW study. In addition, in the UW study, food was provided AL during the day from ~8:00 a.m. to ~4:00 p.m., but was removed during the night. Thus, in both studies, the AL groups were either partially calorie restricted (NIA) or

were under a time-restricted feeding protocol (UW), which should be beneficial. Thus, it would be of considerable interest to explore the effects of time-restricted feeding and circadian phase of CR in future studies in nonhuman primates.

The benefits of circadian alignment of feeding and fasting cycles are widespread across species. It increases life span in flies (*54*), protects against metabolic disorders in mice (*17*, *18*, *55*, *56*), and promotes health in humans (*57*, *58*). We also found that CR promotes widespread metabolic benefits, which include lower body weight with reduced fat content. Furthermore, CR (particularly the longest-lived group, CR-night) attenuated age-related changes observed under AL by improving glucose homeostasis, insulin sensitivity, and hormonal profiles.

Using circadian gene profiling, we found that timing of food intake led to complex genome-wide reprogramming of circadian gene expression in the liver, which is consistent with other findings (*11*, *14*, *33*, *59*–*61*). This emphasizes the importance of considering the sampling time, which is often a snapshot, before concluding whether any intervention increases or decreases the expression of a gene of interest. If a gene has a circadian oscillation, then a single snapshot could lead to the opposite conclusions depending on the time it was taken, particularly in classic CR protocols resembling our CR-day-2h group (*1*, *7*).

Aging was associated with increased expression of the genes involved in immune processes and inflammation and decreased expression of the genes involved in metabolism and circadian biology. CR treatment restored or attenuated many of these age-related changes in gene expression in a manner similar to results reported by the de Cabo (*24*, *34*) and Sassone-Corsi (*14*) groups. Decreases in circadian amplitude of gene expression with age were reduced by CR. Overall, feeding time and fasting had additive effects on CR-mediated life-span extension. Together, CR and time restriction of feeding to the nighttime optimally extended life span and delayed many of the age-related gene expression changes in immune function, inflammation, and metabolism. Further studies are required to determine whether disruption of circadian sleep/wake cycles by daytime feeding also contributes to the day versus night differential responses in the liver transcriptome and life span. Thus, circadian interventions such as timed feeding can enhance the well-known life-span benefits of CR.

There are many links between the circadian system and aging (*20*). High-amplitude circadian rhythms correlate with well-being (*13*, *62*, *63*), whereas clock dysfunction leads to metabolic disorders, premature aging, and reduced life span (*42*, *64*–*67*). During normal aging, rhythms decrease in amplitude and often exhibit a shift in phase (*14*, *20*, *68*–*72*). In both flies and mice, the effects of dietary restriction on life span require core circadian clock genes (*73*–*75*). The circadian system regulates most of the metabolic pathways implicated in longevity (*3*, *4*, *13*, *22*, *32*, *59*, *76*, *77*). Molecules known to function in the regulation of life span by dietary restriction, such as insulin and insulin-like growth factor 1 (IGF-1), Sirtuin1 (SIRT1), nicotinamide phosphoribosyltransferase (NAMPT), AMP-activated protein kinase (AMPK), PPARγ coactivator 1 (PGC-1α), mechanistic target of rapamycin (mTOR), and glycogen synthase kinase 3β (GSK3β), are all intricately involved in the molecular mechanisms of circadian clocks (*77*–*89*). The master circadian transcription factors CLOCK and BMAL1 have direct gene targets in every fundamental metabolic pathway in the liver (*25*, *32*, *90*, *91*). Because of these direct links among the pathways involved in aging and longevity, metabolism, and the circadian clock, our results demonstrate the importance of timing of CR and indicate that optimizing the phase of circadian gene expression may be a powerful intervention for extending life span.

We used C57BL/6J male mice in this study. However, there could be strain- and sex-specific responses worth studying further (*24*, *92*), because, for example, ovarian hormones can protect females against dietary challenges that otherwise disrupt circadian rhythms in males (*55*). In future work, both sexes and multiple genetic backgrounds (*92*–*94*) should be used to explore the broader effects of circadian interventions on aging, and the results may support the application of circadian-timed interventions in human studies.

REFERENCES AND NOTES

1. R. Weindruch, R. L. Walford, S. Fligiel, D. Guthrie, *J. Nutr.* **116**, 641–654 (1986).
2. L. Fontana, L. Partridge, V. D. Longo, *Science* **328**, 321–326 (2010).
3. C. López-Otín, L. Galluzzi, J. M. P. Freije, F. Madeo, G. Kroemer, *Cell* **166**, 802–821 (2016).
4. A. Di Francesco, C. Di Germanio, M. Bernier, R. de Cabo, *Science* **362**, 770–775 (2018).
5. M. Gonzalez-Freire *et al.*, *Ageing Res. Rev.* **59**, 101037 (2020).
6. M. B. Lee, C. M. Hill, A. Bitto, M. Kaeberlein, *Science* **374**, eabe7365 (2021).
7. V. A. Acosta-Rodríguez, M. H. M. de Groot, F. Rijo-Ferreira, C. B. Green, J. S. Takahashi, *Cell Metab.* **26**, 267–277.e2 (2017).
8. S. J. Mitchell *et al.*, *Cell Metab.* **29**, 221–228.e3 (2019).
9. H. H. Pak *et al.*, *Nat. Metab.* **3**, 1327–1341 (2021).
10. F. Damiola *et al.*, *Genes Dev.* **14**, 2950–2961 (2000).
11. C. Vollmers *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21453–21458 (2009).
12. M. Izumo *et al.*, *eLife* **3**, e04617 (2014).
13. O. Froy, *Ageing Res. Rev.* **12**, 931–940 (2013).
14. S. Sato *et al.*, *Cell* **170**, 664–677.e11 (2017).
15. G. Solanas *et al.*, *Cell* **170**, 678–692.e20 (2017).
16. V. D. Longo, S. Panda, *Cell Metab.* **23**, 1048–1059 (2016).
17. D. M. Arble, J. Bass, A. D. Laposky, M. H. Vitaterna, F. W. Turek, *Obesity (Silver Spring)* **17**, 2100–2102 (2009).
18. M. Hatori *et al.*, *Cell Metab.* **15**, 848–860 (2012).
19. A. Chaix, A. Zarrinpar, P. Miu, S. Panda, *Cell Metab.* **20**, 991–1005 (2014).
20. A. Chaudhari, R. Gupta, K. Makwana, R. Kondratov, *Nutr. Healthy Aging* **4**, 101–112 (2017).
21. S. Gill, H. D. Le, G. C. Melkani, S. Panda, *Science* **347**, 1265–1269 (2015).
22. V. A. Acosta-Rodríguez, F. Rijo-Ferreira, C. B. Green, J. S. Takahashi, *Nat. Commun.* **12**, 2862 (2021).
23. M. A. Pellizzon, M. R. Ricci, *Curr. Dev. Nutr.* **4**, nzaa031 (2020).
24. S. J. Mitchell *et al.*, *Cell Metab.* **23**, 1093–1112 (2016).
25. J. Bass, M. A. Lazar, *Science* **354**, 994–999 (2016).
26. M. P. Mattson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16647–16653 (2014).
27. V. S. Valentinuzzi, K. Scarbrough, J. S. Takahashi, F. W. Turek, *Am. J. Physiol.* **273**, R1957–R1964 (1997).
28. A. Turturro, P. Duffy, B. Hass, R. Kodell, R. Hart, *J. Gerontol. A Biol. Sci. Med. Sci.* **57**, B379–B389 (2002).
29. R. C. Northeast *et al.*, *Sleep* **42**, zsz157 (2019).
30. R. Garcia-Valles *et al.*, *Longev. Healthspan* **2**, 14 (2013).
31. A. Andersen, A. Lund, F. K. Knop, T. Vilsbøll, *Nat. Rev. Endocrinol.* **14**, 390–403 (2018).
32. N. Koike *et al.*, *Science* **338**, 349–354 (2012).
33. B. J. Greenwell *et al.*, *Cell Rep.* **27**, 649–657.e5 (2019).
34. M. A. Aon *et al.*, *Cell Metab.* **32**, 100–116.e4 (2020).
35. R. L. Silverstein, M. Febbraio, *Sci. Signal.* **2**, re3 (2009).
36. C. Argmann *et al.*, *PLOS Genet.* **5**, e1000752 (2009).
37. D. Ernszt *et al.*, *Front. Immunol.* **8**, 1515 (2017).
38. Y. Sun *et al.*, *J. Transl. Med.* **17**, 390 (2019).
39. S. P. Liu *et al.*, *Hepatology* **46**, 1413–1425 (2007).
40. M. L. Martínez-Chantar *et al.*, *Hepatology* **47**, 1191–1199 (2008).
41. Y. J. Liao *et al.*, *Proteomics Clin. Appl.* **4**, 394–406 (2010).
42. A. A. Kondratova, R. V. Kondratov, *Nat. Rev. Neurosci.* **13**, 325–335 (2012).
43. Z. Xiang *et al.*, *J. Exp. Clin. Cancer Res.* **38**, 314 (2019).
44. W. Abi Habib *et al.*, *Genet. Med.* **20**, 250–258 (2018).
45. R. A. Tumasian3rd *et al.*, *Nat. Commun.* **12**, 2014 (2021).
46. M. Xu *et al.*, *Onco Targets Ther.* **9**, 4023–4035 (2016).
47. R. Allada, J. Bass, *N. Engl. J. Med.* **384**, 550–561 (2021).
48. F. Rijo-Ferreira, J. S. Takahashi, *Genome Med.* **11**, 82 (2019).
49. M. A. Pellizzon, M. R. Ricci, *Curr. Protoc. Toxicol.* **77**, e55 (2018).
50. M. A. Pellizzon, M. R. Ricci, *Nutr. Metab. (Lond.)* **15**, 3 (2018).
51. R. J. Colman *et al.*, *Nat. Commun.* **5**, 3557 (2014).
52. J. A. Mattison *et al.*, *Nat. Commun.* **8**, 14063 (2017).
53. J. A. Mattison *et al.*, *Nature* **489**, 318–321 (2012).
54. M. Ulgherait *et al.*, *Nature* **598**, 353–358 (2021).
55. B. T. Palmisano, J. M. Stafford, J. S. Pendergast, *Front. Endocrinol. (Lausanne)* **8**, 44 (2017).
56. A. Chaix, E. N. C. Manoogian, G. C. Melkani, S. Panda, *Annu. Rev. Nutr.* **39**, 291–315 (2019).
57. M. J. Wilkinson *et al.*, *Cell Metab.* **31**, 92–104.e5 (2020).
58. E. N. C. Manoogian, A. Chaix, S. Panda, *J. Biol. Rhythms* **34**, 579–581 (2019).
59. G. Asher, P. Sassone-Corsi, *Cell* **161**, 84–92 (2015).
60. K. L. Eckel-Mahan *et al.*, *Cell* **155**, 1464–1478 (2013).
61. G. Manella *et al.*, *Nat. Metab.* **3**, 829–842 (2021).
62. B. He *et al.*, *Cell Metab.* **23**, 610–621 (2016).
63. O. Froy, *Proc. Nutr. Soc.* **77**, 216–222 (2018).
64. R. V. Kondratov, A. A. Kondratova, V. Y. Gorbacheva, O. V. Vykhovanets, M. P. Antoch, *Genes Dev.* **20**, 1868–1873 (2006).
65. K. A. Lamia, K. F. Storch, C. J. Weitz, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15172–15177 (2008).
66. B. Marcheva *et al.*, *Nature* **466**, 627–631 (2010).
67. Y. V. Dubrovsky, W. E. Samsa, R. V. Kondratov, *Aging (Albany NY)* **2**, 936–944 (2010).
68. K. Scarbrough, S. Losee-Olson, E. P. Wallen, F. W. Turek, *Am. J. Physiol.* **272**, R1219–R1225 (1997).
69. S. Yamazaki *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10801–10806 (2002).
70. M. A. Hofman, D. F. Swaab, *Ageing Res. Rev.* **5**, 33–51 (2006).
71. E. M. Gibson, W. P. Williams 3rd, L. J. Kriegsfeld, *Exp. Gerontol.* **44**, 51–56 (2009).
72. O. Froy, R. Miskin, *Aging (Albany NY)* **2**, 7–27 (2010).
73. S. D. Katewa *et al.*, *Cell Metab.* **23**, 143–154 (2016).
74. S. A. Patel, A. Chaudhari, R. Gupta, N. Velingkaar, R. V. Kondratov, *FASEB J.* **30**, 1634–1642 (2016).
75. M. Ulgherait *et al.*, *Nat. Commun.* **11**, 1927 (2020).
76. S. S. Fonseca Costa, J. A. Ripperger, *Front. Neurol.* **6**, 43 (2015).
77. D. C. Levine *et al.*, *Mol. Cell* **78**, 835–849.e7 (2020).
78. J. Bass, J. S. Takahashi, *Science* **330**, 1349–1354 (2010).
79. M. M. Bellet, R. Orozco-Solis, S. Sahar, K. Eckel-Mahan, P. Sassone-Corsi, *Cold Spring Harb. Symp. Quant. Biol.* **76**, 31–38 (2011).
80. G. Asher, U. Schibler, *Cell Metab.* **13**, 125–137 (2011).
81. E. E. Zhang *et al.*, *Cell* **139**, 199–210 (2009).
82. G. Asher *et al.*, *Cell* **134**, 317–328 (2008).

83. C. Liu, S. Li, T. Liu, J. Borjigin, J. D. Lin, *Nature* **447**, 477–481 (2007).
84. Y. Harada, M. Sakai, N. Kurabayashi, T. Hirota, Y. Fukada, *J. Biol. Chem.* **280**, 31714–31721 (2005).
85. T. Hirota *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20746–20751 (2008).
86. K. A. Lamia *et al*., *Science* **326**, 437–440 (2009).
87. Y. Nakahata *et al*., *Cell* **134**, 329–340 (2008).
88. Y. Nakahata, S. Sahar, G. Astarita, M. Kaluzova, P. Sassone-Corsi, *Science* **324**, 654–657 (2009).
89. K. M. Ramsey *et al*., *Science* **324**, 651–654 (2009).
90. S. Panda, *Science* **354**, 1008–1015 (2016).
91. H. Reinke, G. Asher, *Nat. Rev. Mol. Cell Biol.* **20**, 227–241 (2019).
92. R. Yuan *et al*., *Aging Cell* **8**, 277–287 (2009).
93. R. Korstanje, L. L. Peters, L. L. Robinson, S. D. Krasinski, G. A. Churchill, *Geroscience* **43**, 2129–2137 (2021).
94. C. Selman, W. R. Swindell, *EMBO J.* **37**, e100862 (2018).

ACKNOWLEDGMENTS

We thank G. Martinez, C. Joseph, K. Brown, and D. Bassowou for assistance with animal care and maintenance; I. Kornblum for preparing mRNA-seq libraries; G. Kilaru for mapping NGS files; and S. Yamazaki for helpful discussions. We apologize for the omission of relevant citations because of space constraints. **Funding:** This work was supported by the Howard Hughes Medical Institute (J.S.T.); the National Institutes of Health (NIH) National Institute on Aging (grant R01 AG045795 and R56 AG072736 to J.S.T. and C.B.G.); the NIH National Institute of General Medical Sciences (NIGMS) (grant R35 GM127122 to C.B.G. and grant K99 GM132557 to F.R.-F.); and the Milky Way Research Foundation (grant MWRF210823 to J.S.T. and C.B.G.). J.S.T. is an investigator and F.R.-F. was an associate at the Howard Hughes Medical Institute. **Author contributions:** Conceptualization: J.S.T., C.B.G., V.A.R.; Funding acquisition: J.S.T., C.B.G.; Investigation: V.A.R., F.R.-F., M.I., P.X., M.W.-C.; Methodology: V.A.R., J.S.T.; Project administration: J.S.T., C.B.G.; Supervision: J.S.T., C.B.G., V.A.R.; Visualization: V.A.R., F.R.-F.; Writing – original draft: V.A.R., J.S.T., C.B.G., F.R.-F.; Writing – review and editing: V.A.R., J.S.T., C.B.G., F.R.-F. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data are available in the main text or the supplementary materials. RNA-seq data have been deposited in the Gene Expression Omnibus (accession no. GSE190939). **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.abk0297
Materials and Methods
Figs. S1 to S14
References (*95*–*106*)
Data S1 to S6
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 17 June 2021; accepted 7 April 2022
Published online 5 May 2022
10.1126/science.abk0297

MARINE VIROME

Diversity and ecological footprint of Global Ocean RNA viruses

Guillermo Dominguez-Huerta[1,2,3]†, Ahmed A. Zayed[1,2,3]†, James M. Wainaina[1,3], Jiarong Guo[1,2,3], Funing Tian[1,3], Akbar Adjie Pratama[1,2], Benjamin Bolduc[1,2,3], Mohamed Mohssen[1,3,4], Olivier Zablocki[1,2,3], Eric Pelletier[5,6], Erwan Delage[6,7], Adriana Alberti[5,6]‡, Jean-Marc Aury[5], Quentin Carradec[5,6], Corinne da Silva[5], Karine Labadie[5,6], Julie Poulain[5,6], Tara Oceans Coordinators§, Chris Bowler[6,8], Damien Eveillard[6,7], Lionel Guidi[6,9], Eric Karsenti[6,8,10], Jens H. Kuhn[11], Hiroyuki Ogata[12], Patrick Wincker[5,6], Alexander Culley[13], Samuel Chaffron[6,7], Matthew B. Sullivan[1,2,3,4,14]*

DNA viruses are increasingly recognized as influencing marine microbes and microbe-mediated biogeochemical cycling. However, little is known about global marine RNA virus diversity, ecology, and ecosystem roles. In this study, we uncover patterns and predictors of marine RNA virus community- and "species"-level diversity and contextualize their ecological impacts from pole to pole. Our analyses revealed four ecological zones, latitudinal and depth diversity patterns, and environmental correlates for RNA viruses. Our findings only partially parallel those of cosampled plankton and show unexpectedly high polar ecological interactions. The influence of RNA viruses on ecosystems appears to be large, as predicted hosts are ecologically important. Moreover, the occurrence of auxiliary metabolic genes indicates that RNA viruses cause reprogramming of diverse host metabolisms, including photosynthesis and carbon cycling, and that RNA virus abundances predict ocean carbon export.

The Global Ocean is dominated by plankton communities that are essential to sustain life on Earth. Plankton are at the base of the food web for marine and terrestrial organisms and drive planetary biogeochemical cycles (*1*, *2*). Because nearly half of Earth's primary production derives from ocean plankton, carbon cycling and biodiversity studies have long been a focus in oceanography (*3*). In addition, marine plankton are central to the biological carbon pump because their activity determines whether dissolved carbon dioxide is assimilated into biomass that can be sequestered to the deep ocean or recycled in surface waters and likely released to the atmosphere (*4*, *5*). Thus, understanding ocean biodiversity, carbon export, and related chemical transformations is critical to predicting the changing role of the ocean in the Anthropocene.

Plankton are susceptible to virus infection. Double-stranded DNA (dsDNA) viruses have been increasingly recognized as major ecosystem players (*6*), whereas RNA viruses have been less well-studied owing to methodological challenges (*7*). It is clear, however, that marine RNA viruses are likely important in marine ecosystems, as they (i) are abundant (*8*, *9*), (ii) infect protists and invertebrates that are central to ocean biogeochemical cycling (*10*), and (iii) have been statistically associated with termination of algal blooms (*11*, *12*) and modulation of host diversity (*13*). Despite literature increasingly presenting RNA viruses as a likely major force behind biogeochemistry (*6*, *14*, *15*), empirical data are challenging to obtain. Recent sequencing surveys, including from the oceans, have identified thousands of previously unknown RNA viruses that constitute genus- or subfamily-rank taxa (*16*–*18*) as well as phylum-rank taxa (*19*). However, research on the ecology of RNA viruses has been limited to small spatial scales among pelagic waters and/or viruses associated with larger plankton of a few species (table S1). This lack of ecological context, particularly over large scales, limits the incorporation of RNA viruses into predictive models.

Previously, we analyzed 771 metatranscriptomes (provided by *Tara* Oceans Expeditions) that span diverse ocean waters, depths, organismal size fractions, and sequencing library approaches (Fig. 1A, fig. S1, table S2 for sample metadata, and materials and methods) to identify and quantify RNA viruses (*19*). This effort led to the identification of 44,779 RNA virus contigs that were dereplicated to 5504 "species"-level virus operational taxonomic units (vOTUs), for which we established taxonomy, evolutionary origins, and biogeography. In this work, we leverage these data to generate and test several existing hypotheses about RNA virus diversity and their ecological roles throughout the Global Ocean.

RNA virus meta-community analyses reveal distinct ecological zones

Given the importance of marine plankton (*2*), scientists have long sought to understand their ecological patterns and drivers through space and/or time. Temporal studies have revealed seasonal-, depth-, and nutrient-related local or regional drivers of plankton species diversity and community composition, whereas systematic surveys sought to examine these ecological patterns and drivers on a global scale (table S3). However, none of these global studies included RNA viruses. Hence, we used our previously generated RNA vOTUs (*19*), preclustered at 90% average nucleotide identity across 80% of the shorter sequence length and 1-kb minimum contig length (materials and methods), and their relative abundances, estimated by means of metatranscriptomic read mapping (materials and methods), to investigate marine RNA virus ecology globally.

By using a statistical method that nonlinearly deconvolutes high-dimensional data into two-dimensional space (Fig. 1B; *t*-distributed stochastic neighbor embedding, fig. S2, A to C) and classical hierarchical clustering techniques (fig. S2D) on Bray-Curtis dissimilarity matrices of RNA vOTU relative abundances (materials and methods), we show that Global Ocean RNA virus communities can be assigned to four ecological zones: Arctic, Antarctic, Temperate and Tropical Epipelagic, and Temperate and Tropical Mesopelagic. This classification into only four ecological zones contrasts with the 56 biogeochemical provinces that are classically described for the surface oceans, where nutrients and primary productivity drive plankton community composition (*20*). However, the four ecological zone assignments are nearly identical (115 of 118 shared samples) to those

[1]Department of Microbiology, The Ohio State University, Columbus, OH 43210, USA. [2]EMERGE Biology Integration Institute, The Ohio State University, Columbus, OH 43210, USA. [3]Center of Microbiome Science, The Ohio State University, Columbus, OH 43210, USA. [4]The Interdisciplinary Biophysics Graduate Program, The Ohio State University, Columbus, Ohio 43210, USA. [5]Génomique Métabolique, Genoscope, Institut François-Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91000 Evry, France. [6]Research Federation for the Study of Global Ocean Systems Ecology and Evolution, FR2022/*Tara* Oceans GOSEE, 75016 Paris, France. [7]Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France. [8]Institut de Biologie de l'Ecole Normale Supérieure, Ecole Normale Supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France. [9]Sorbonne Université, CNRS, Laboratoire d'Océanographie de Villefanche, LOV, F-06230 Villefranche-sur-mer, France. [10]Directors' Research European Molecular Biology Laboratory, 69117 Heidelberg, Germany. [11]Integrated Research Facility at Fort Detrick, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Fort Detrick, Frederick, MD 21702, USA. [12]Institute for Chemical Research, Kyoto University, Kyoto 611-0011, Japan. [13]Département de Biochimie, Microbiologie et Bio-informatique, Université Laval, Québec, QC G1V 0A6, Canada. [14]Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH 43210, USA.
*Corresponding author. Email: sullivan.948@osu.edu
†These authors contributed equally to this work. ‡Present address: Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198 Gif-sur-Yvette, France. §The *Tara* Oceans Coordinators are listed in the Supplementary Materials.

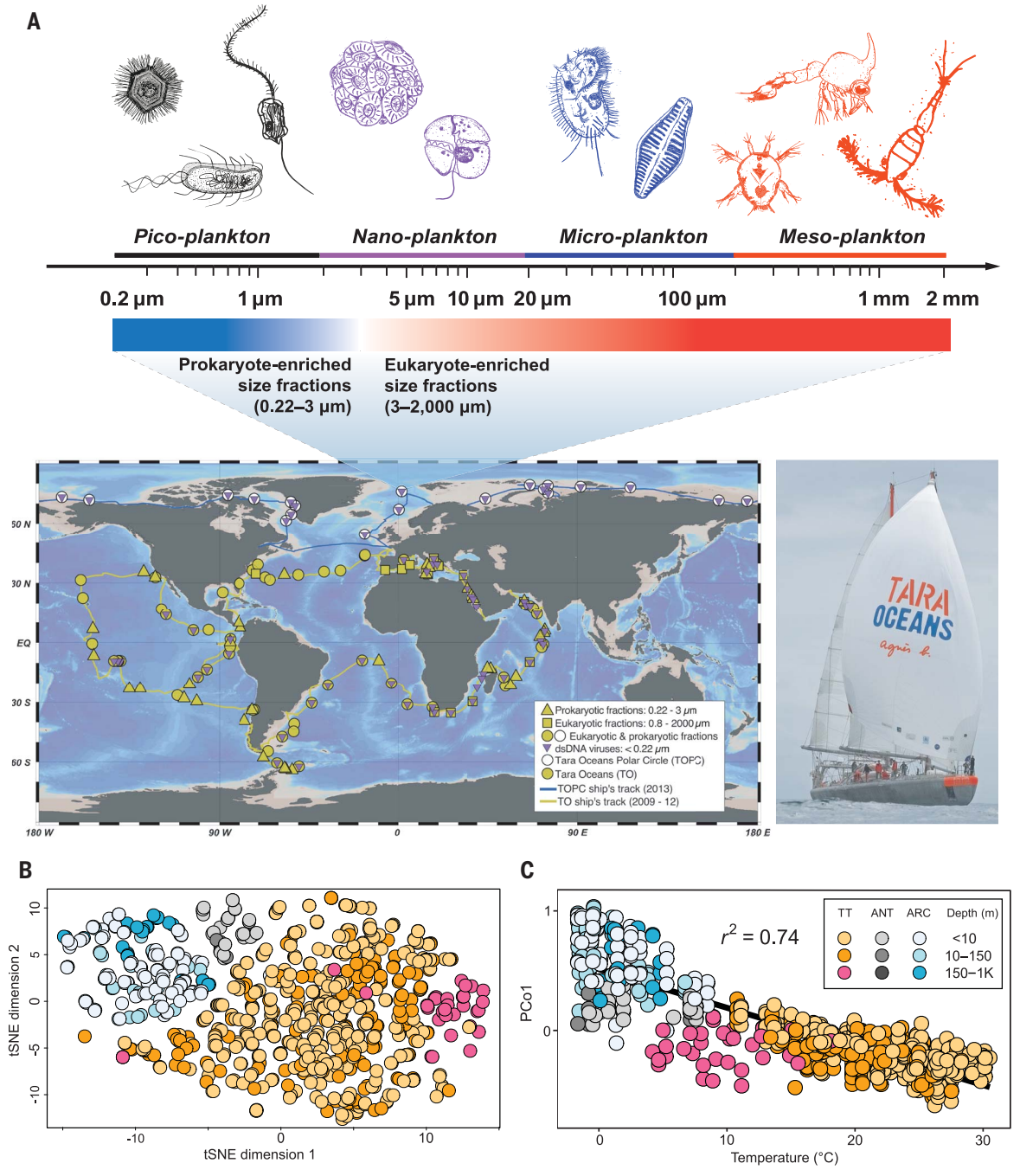


Fig. 1. The cross-domain Global Ocean plankton sampling and resultant RNA virus meta-communities identified from the metatranscriptomes. (**A**) Global Ocean sampling map shows the cruise of the *Tara* Oceans and *Tara* Oceans Polar Circle expeditions and the location of their stations, which are shown with green and white shapes, respectively. Down-pointing triangles indicate stations from where dsDNA viromes were previously collected. Up-pointing triangles, squares, and circles show stations with samples of prokaryote-enriched size fractions, eukaryote-enriched size fractions, and both, respectively. The upper blowout panel shows a graded arrow that represents a logarithmic scale of the plankton organismal size fractions captured in this study. The four operational size fractions (piconanoplankton, nanoplankton, microplankton, and mesoplankton) are indicated by the top colored bars and are classified as "prokaryote-enriched" or "eukaryote-enriched" size fractions (highlighted by the bottom gradient-colored bars). Such categories, despite being enriched in a type of organism, do not exclude other types. Thus, prokaryote-enriched samples could contain giant viruses and picoeukaryotes, and eukaryotic holobionts of eukaryote-enriched samples could harbor prokaryotes or viruses either as symbionts or food. A picture of the research vessel *Tara* is included as well. (**B**) Statistical analysis [*t*-distributed stochastic neighbor embedding (*t*-SNE)] of a Bray-Curtis dissimilarity matrix that was calculated from all RNA virus sequence samples in this study regardless of size fraction or library preparation method. Dot colors follow the legend shown in (C) (also see figs. S4 and S5 for vOTU definition sensitivity analyses). (**C**) Regression analysis of the first coordinate of a principal coordinate analysis (PCo1) of the same Bray-Curtis dissimilarity matrix in (A) (also see fig. S2) and temperature, which shows that samples across all the size fractions were separated by their local temperatures with an r^2 of 0.74 (*P* values = 0). ANT, Antarctic; ARC, Arctic; TT, Temperate and Tropical.

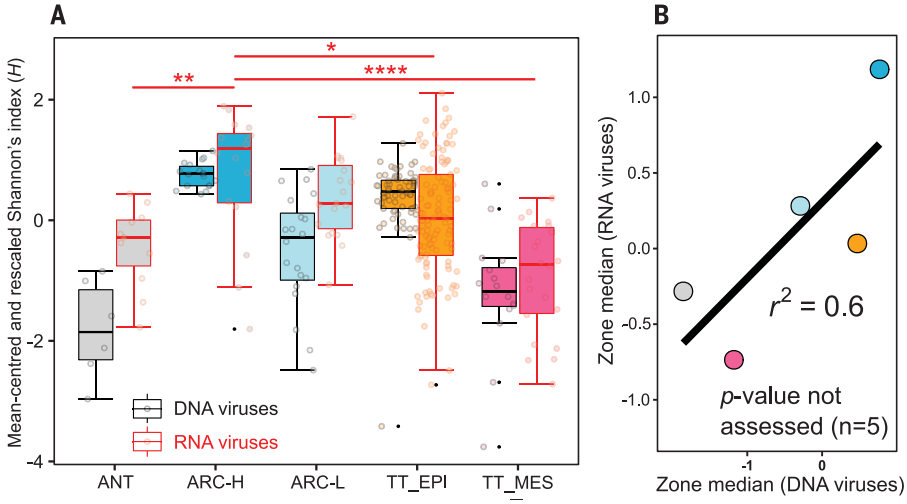


Fig. 2. RNA and DNA virus "species"-level diversity show large-scale congruence. (**A** and **B**) Boxplot (A) and regression (B) analyses of RNA and DNA virus "species"-level diversity across their shared ecological zones. Shannon's *H* values were mean-centered and rescaled across the two virus nucleic acid types for visual comparisons. All boxplots show medians and quartiles. The medians of each boxplot were used for direct regression analysis. Statistical support (Tukey honest significant differences method on an analysis of variance) is indicated in the figure as follows: ∗adjusted $P < 0.05$, ∗∗adjusted $P < 0.01$, and ∗∗∗∗adjusted $P < 0.000001$. Only RNA viruses from the prokaryotic fraction were used (see fig. S3 for comparison with the eukaryotic fractions) as this fraction showed the smallest library preparation biases (fig. S1 and materials and methods). ANT, Antarctic; ARC-H, Arctic high diversity; ARC-L, Arctic low diversity; TT_EPI, Temperate and Tropical Epipelagic; TT_MES, Temperate and Tropical Mesopelagic.

that were inferred for prokaryotic dsDNA viruses (materials and methods; the fifth Bathypelagic zone that was inferred from dsDNA virus analyses was not sampled here) (*21*) and largely parallel to those from broader *Tara* Oceans Consortium work on prokaryotes (*22*). Before this study, these ecological zone analyses had not been performed for eukaryotes or eukaryotic RNA viruses. Also previously, transport or migration of eukaryotic plankton across ocean surface biomes and layers was thought to erode the boundaries between these ecological zones (*23*). Our and other recent eukaryotic data (*24*) challenge this hypothesis.

Investigation of ecological parameters that potentially drive community structure at large scale revealed that temperature alone could explain most RNA virus community composition variation along the first ordination axis (Fig. 1C). Other ecological drivers, including oxygen, depth, and nutrient availability, may shape plankton community composition (table S3, A10 to A14), but these often co-vary with temperature. Limited sampling in these previous, geographically constrained studies led to the hypothesis that depth is the main driver of plankton community composition. With global data now available, it is apparent that temperature variance potentially drives stratification in nonpolar regions (fig. S2, E and F) and selects for cold-adapted communities in polar regions. A temperature-driven RNA virus community composition complements that for dsDNA viruses (*21*), prokaryotes (*22*), eukaryotes (*24*), and their interactions (*25*).

Differential predictors of RNA virus global and local "species"-level diversity

Comparison of the diversity patterns of RNA (this study) and dsDNA (*21*) viruses revealed highly concordant large-scale patterns, including previously identified (*21*) high- and low-diversity regions of the Arctic Ocean (ARC-H and ARC-L) (Fig. 2). However, local diversity comparisons (i.e., per-sample comparisons) showed that the concordance, despite being significant ($P < 0.02$), was modest ($r \approx 0.25$ per each Pearson's and Spearman's tests), which suggests that small-scale diversity drivers may differ for DNA and RNA viruses. When examining the large suite of environmental variables available for our samples (table S4) for possible correlations with RNA and dsDNA virus diversity, we accounted for collinearity using a systems biology network analysis framework to reduce environmental factor dimensionality into fewer environmental "modules" (Fig. 3 and materials and methods).

We found, first, that similar to dsDNA viruses (*21*), temperature (cyan module in Fig. 3) was not the best predictor of RNA virus diversity. Instead, nutrients (white module in Fig. 3) were prominent predictors of species diversity for both dsDNA and RNA viruses, along with other signatures of primary productivity (violet module in Fig. 3). Second, in our previous study on dsDNA viruses (*21*), we showed that the link between dsDNA virus diversity and nutrients might be through primary productivity, because photosynthetic coccolithophores' abundance and particulate inorganic carbon (PIC) concentration covaries with dsDNA virus diversity (light green module in Fig. 3). More recently, the relationship between dsDNA viruses and PIC has been posited to be abiotic on the basis of direct virus-mediated mineral precipitation (*26*). Unlike dsDNA virus diversity, RNA virus diversity does not correlate with the PIC module but does still correlate with primary productivity pigment concentrations such as chlorophyll b (yellow module in Fig. 3), which indicates, as expected, that dsDNA and RNA viruses infect different hosts. This and other biological features of RNA viruses, such as their shorter and faster-evolving genomes, higher burst sizes, lytic lifestyles, and eukaryotic hosts, are hypothesized to drive virus–host interaction and ecosystem impact differences from dsDNA viruses (*27*). Models that are based on known RNA virus biological features also lend support to this idea (*6*, *7*, *27*, *28*). We interpret the small-scale differences in diversity patterns, despite high concordance at the large scale, as also deriving from varied biological features across RNA and dsDNA viruses.

Together, these findings indicate that the underlying large-scale potential drivers for virus community composition (which encompasses the identity and abundance of vOTUs) and species diversity (which encompasses the vOTUs' richness and distribution evenness) act similarly for the RNA viruses of eukaryotes and the dsDNA viruses of prokaryotes. For virus community composition, perhaps this is not surprising, given that likely host community compositions (planktonic prokaryotes and microbial eukaryotes) also appear to be mainly driven by temperature (*22*, *24*, *29*). For virus diversity, the relationship with host diversity can be more complex (see "RNA virus 'species'-level diversity along ecological gradients"). Locally, the varying biological features of RNA viruses are hypothesized (*7*, *28*) to drive virus–host interaction and ecosystem impact differences between largely prokaryotic dsDNA viruses and eukaryotic RNA viruses. For local diversity predictors, our findings are consistent with this hypothesis.

RNA virus "species"-level diversity along ecological gradients

The physicochemical tolerances, or ecological gradients, of RNA viruses are not understood. Organismal diversity typically decreases with

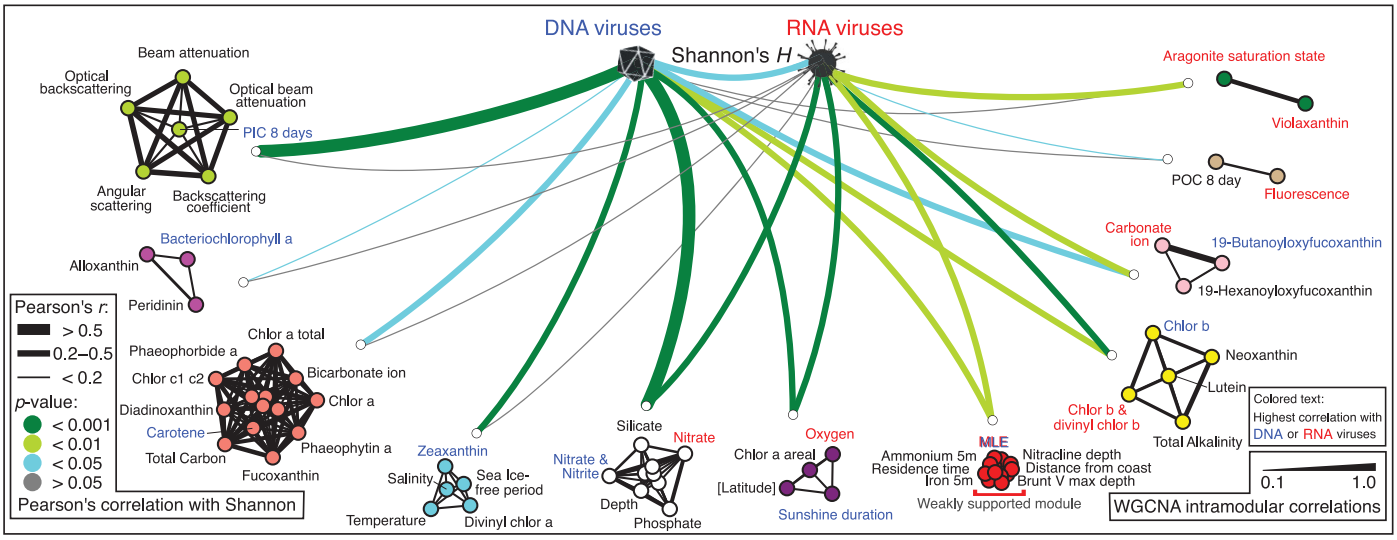


Fig. 3. "Species"-level diversity correlates of marine RNA viruses. Weighted gene correlation network analysis (WGCNA)–supported modules (to account for collinearity) of environmental variables (materials and methods) showing the cofactors of RNA and DNA virus diversity. Modules are Pearson correlated to the Shannon's *H* values of each virus group. Shown are only those relationships that were statistically supported by both Pearson's and Spearman's tests. Only RNA viruses from the prokaryotic fraction were used (see Fig. 2 for explanation). Notably, aragonite and carbonates could be indicative of coccolithophores, whereas violaxanthin and the latitude-chlor a signal could be related to diatoms. MLE, maximum Lyapunov exponent; POC, particulate organic carbon.

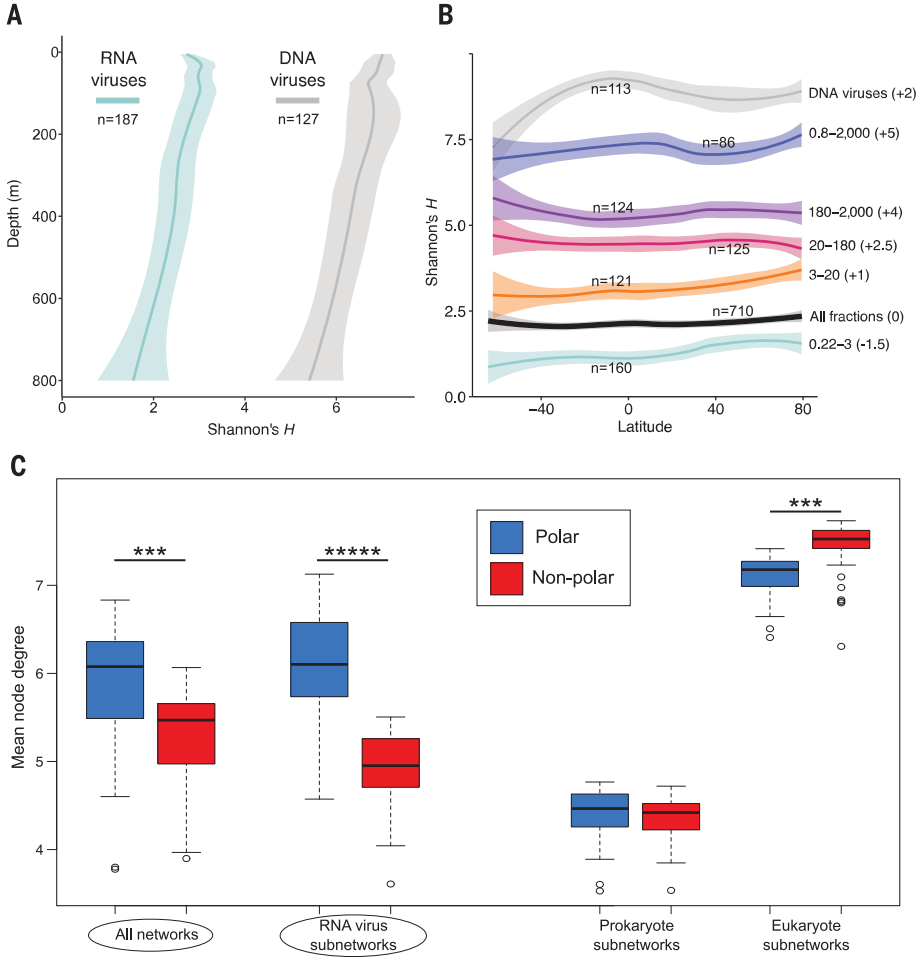


Fig. 4. RNA virus "species"-level diversity across depth and latitudinal gradients. (**A**) Locally estimated scatterplot smoothing (LOESS) smooth plots showing the depth distributions of "species"-level diversity for RNA and DNA viruses. Shown are only RNA viruses from the prokaryotic fraction because of the very limited number of deep ocean samples from the eukaryotic fraction (eukaryotic fraction results are shown in fig. S3). (**B**) LOESS plots showing the latitudinal distributions of "species"-level diversity for DNA (gray) and RNA viruses (remaining colors). Plots are nudged along the *y*-axis (with a baseline offset as indicated in the parentheses on the right) for visibility. Size fraction and nudge value are indicated next to each plot, with the collective estimate of Shannon's *H* values across all the size fractions of RNA viruses shown in black. On all of the smoothing plots, the lines represent the LOESS best fit for the samples included (*n*), whereas the lighter band corresponds to the 95% confidence interval of the fit (also see figs. S4 and S5 for vOTU definition sensitivity analyses). (**C**) Global and organismal domain-specific co-occurrence networks connectivity (mean node degree) in polar versus nonpolar samples showing that the significantly higher connectivity in the polar waters (ellipses) is driven solely by RNA viruses. All boxplots show medians and quartiles. Statistical significance was assayed by the Mann-Whitney *U* test and is documented in the figure as follows: ∗adjusted $P < 0.05$, ∗∗∗adjusted $P < 0.0001$, and ∗∗∗∗∗adjusted $P < 0.0000001$.

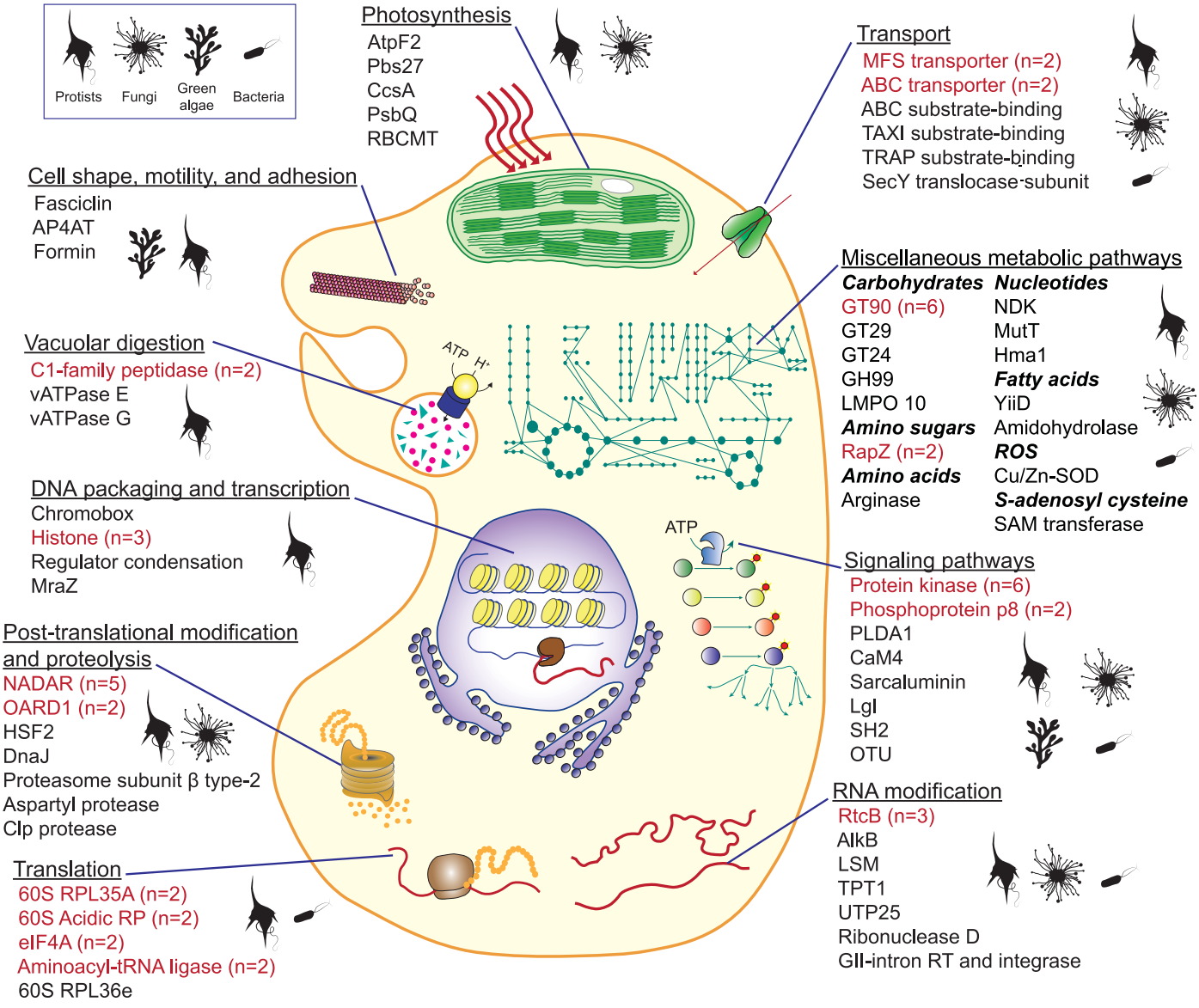


Fig. 5. Functional diversity of AMGs carried by marine RNA viruses. Schematic representation of the hypothesized roles played in manipulation of host metabolism by RNA virus AMGs, which are separated according to functional categories. Red text corresponds to proteins that were found encoded independently in several vOTUs with the number of vOTUs listed in parentheses. The putative hosts, which were inferred by using available information for RNA viruses with established orthornaviran taxonomy, are indicated by organism silhouettes in each section. Inferred plants were interpreted as their closest relatives, chlorophytes (green algae), in the marine environment. Bacteria were inferred from picobirnavirids. Annotated proteins associated with multiple, disparate cellular processes or whose function remains obscure are not shown (see annotation details for corresponding vOTUs and virus contigs in table S6). ABC, ATP-binding cassette; TRAP, tripartite ATP-independent periplasmic.

depth (*30*), as does dsDNA virus diversity (*21*), and we found that RNA virus diversity also decreases with depth (Fig. 4A and fig. S3). Latitudinal diversity gradients are characterized by relatively low polar and high equatorial diversity for most terrestrial flora and fauna (*31*, *32*) and oceanic plankton (*33*). However, paradoxically, prokaryotic dsDNA virus diversity tends to increase in the Arctic (*21*, *33*), unlike their hosts' diversity (*34*, *35*). Thus, to establish baseline paradigms for RNA viruses, we assessed how RNA virus diversity varied with latitude and how it compares with eukaryotic diversity across our Global Ocean dataset. This revealed no obvious latitudinal pattern for RNA virus diversity, regardless of the size fraction (Fig. 4B and fig. S3; also see figs. S4 and S5 for other sensitivity analyses), which is reminiscent of the deviation seen for dsDNA viruses (*21*). This disconnect of virus and host diversity also has a precedent among nonviruses [see eukaryotic photosynthetic intracellular symbionts and their eukaryotic hosts (*33*)]. We hypothesize that this disconnect is caused by the differential impacts of temperature, allowing (i) viral particles to be better preserved in cold temperatures and/or (ii) more viruses of distinct species to interact with the same host organism in polar waters. The former hypothesis has some support in literature (*36*), whereas the latter is untested.

To test the latter hypothesis, we built an abundance-based co-occurrence network that integrated RNA viruses, prokaryotes, and eukaryotes (materials and methods) to predict hosts for these RNA viruses [sensu (*25*)]. Assuming that the overall topology of the network is relatively representative, even if specific predictions are not accurate (see the predicted

hosts section below), we compared the average number of connections per taxon (i.e., mean degree) in polar and nonpolar samples. This comparison showed significantly more connections in polar samples than nonpolar samples, and this feature was solely driven by RNA viruses (Fig. 4C). This result was unexpected but is in line with a recent ecological network theory prediction that used data from 511 mammal-infecting viruses to show a nonlinear relationship between host and virus diversity (*37*), which was interpreted to be a result of host sharing among different sets of viruses of separate species.

Hence, although the ecological zones and potential ecological drivers of marine RNA viruses (Fig. 1, B and C) and their expected eukaryotic hosts (*24*) were similar in our datasets, the species diversity relationships of RNA viruses and their hosts can be more complex on a global scale.

Marine RNA viruses and inferred local and global ecological impact

First, we sought to place RNA virus diversity data into an ecosystem context by assessing local- to global-scale impacts by means of infected plankton hosts or altered metabolisms (local scale) versus systems-level ecosystem impact (global scale). We predicted hosts for our vOTUs using three approaches: (i) host information available for viruses of established taxa, (ii) abundance-based co-occurrence, and/or (iii) endogenous virus element (EVE) signatures (fig. S6). Although these results provide only broad taxon rank host predictions, as in silico host inferences for RNA viruses are not well-established, they indicated infection of diverse organisms of ecological interest, predominantly protists and fungi, and, to a lesser extent, invertebrate metazoans (table S5). We also explored alternative eukaryotic genetic codes for host prediction, which revealed 11 known alternative, eukaryotic genetic codes in 6.8% of the vOTUs and indicated microbial eukaryotes (including mitochondria of yeast, mold, protozoans, and chlorophyceans and nuclear codes of several ciliates) and metazoans (mitochondria of invertebrates) as putative hosts (table S5). Notably, these inferred hosts are associated with diverse ecological functions, including phototrophy (e.g., bacillariophytes), phagotrophy (e.g., ciliates), mixotrophy (e.g., dinophyceaens), saprotrophy (e.g., ascomycetes), parasitism (e.g., alveolates), grazing (e.g., arthropods), and filter feeding (e.g., annelids). Furthermore, several of these hosts, including certain invertebrate metazoans and particularly protists and fungi, are also recognized as critical contributors to the biological carbon pump. Although host prediction is challenging, these findings add support to prior work at smaller scales (table S1) that indicate that RNA viruses are central ecological players in the oceans. These findings also indicate that, although prokaryotic cells outnumber eukaryotic organisms in the oceans, few RNA viruses (only 3.4% of the vOTUs) infect bacteria, a result that is consistent with previous marine virome and virus isolate reports (*7*).

Second, ecosystem impact might be inferred from the "cellular" protein sequences that we identified in the RNA virus genomes, which we posited may parallel the "auxiliary metabolic genes" (AMGs) that are ecologically important in marine prokaryotic dsDNA viruses (*38*). Although such cellular protein sequences are uncommon in RNA virus genomes, either as independent open reading frames or as parts of larger virus proteins, we found 72 functionally distinct AMGs in 95 vOTUs (table S6). Together, these may hint at how RNA viruses manipulate host physiology to maximize virus production (Fig. 5). Although chimeric assemblies might artifactually link AMGs to virus RNA–directed RNA polymerases (RdRP) sequences, several lines of evidence argue against this possibility: (i) 15 AMG–RdRP linkages were observed at multiple sampling sites (Fig. 5), and (ii) even though RNA viruses are rarely represented in metatranscriptomes (*16*), long-read sequencing captured three AMG–RdRP linkages (data S1). In addition, no AMGs were present in any of the 14 virus contigs that were putatively derived from EVEs (data S2 and materials and methods). Mechanistically, we presume such AMGs were acquired by RNA virus genomes through copy-choice recombination with cellular RNAs, as was originally suggested for ubiquitin in togaviruses (*39*). We identified 12 previously reported cases of such RdRP-linked AMGs, but only three studies assessed their functional context in virus infection (table S6). Thus, we used this larger dataset to explore the possible biology that such AMGs might offer to RNA viruses and ecosystems.

Functionally, the 72 AMG types were diverse, with only four cases overlapping with the 12 previously reported AMGs in RNA virus genomes (table S6 and data S1). The most common functional type of AMG (15.8%) was involved in RNA modifications (RtcB, AlkB, and RNA 2′-phosphotransferase) and posttranslational modifications (NADAR and OARD1), which may reflect the common need of viruses to evade host antiviral responses through the repair of virus RNAs and proteins (*40*, *41*). Given that viruses must reprogram cells toward virus progeny production and that RNA viruses have relatively short genomes, it was not surprising to see that protein kinases were abundant (14.8%), as they would allow broad reprogramming capability through limited genetic capacity. The frequency of AMGs suggested that a suite of other processes are affected by marine RNA viruses, including carbohydrate metabolism (10.9%), translation (8.9%), nutrient transport (7.9%), photosynthesis (5.9%), and vacuolar digestion (4.0%). We posit that many of these AMGs represent ocean-specific RNA virus adaptations that help cellular "virus factories" maximize output in the often ultralimiting nutrient conditions of seawater.

Recent experimental work has emerged to assess how DNA viruses affect ocean carbon export over small scales (*42*, *43*). We sought to complement these efforts through Global Ocean assessment of RNA viruses by using previously developed machine learning and ecosystem modeling approaches (materials and methods) (*10*) to evaluate in silico whether RNA viruses might affect ocean carbon export. This revealed that RNA virus abundances were strongly predictive of ocean carbon flux and identified specific vOTUs that were most significant for these predictions (fig. S7 and table S7). Specifically, from 5504 vOTUs, 1,243 were identified as part of four highly significant subnetworks (P values ≈ 0) of RNA viruses that strongly predicted carbon flux variation (fig. S7A). Notably, subnetwork-specific topology interrogation by partial least-squares regression modeling and leave-one-out cross-validation techniques (materials and methods) showed that these subnetworks represent predictive community biomarkers for carbon export (cross-validated r^2 up to 0.79, and, critically, in a 1:1 ratio, which implies capturing the correct magnitude in the models) (fig. S7A). Further, these techniques very conservatively identified 11 RNA viruses that were most predictive of carbon flux (i.e., VIP score) (table S7 and fig. S7B) and offer ideal targets for follow-on hypothesis testing. Chlorophytes and haptophytes could be assigned as hosts for two of these viruses (fig. S7B). These algal hosts are thought to be critical components in the biological carbon pump (table S3, A17 to A19).

Conclusions

For decades, extensive studies have focused on plankton dynamics and activity to infer the pairwise links among plankton and carbon export, including recent experimental work with viruses (*42*, *43*). Because these seminal studies were focused on narrow geographic ranges or oceanic provinces, we sought here to instead explore Global Ocean signals by taking advantage of the uniform *Tara* Oceans strategy for sampling plankton and sinking particles to broadly investigate oceanic conditions and ecosystem biota (*10*). Hence, although limited by single time points or "snapshot" sampling, combining these measurements with a robust statistical framework (i.e., network-based, cross-validated, multivariate-aware correlation analysis) enables statistical exploration to establish hypotheses about key ecosystem players. For this, we can leverage the context of hypothesized interactions (*25*) instead of using the more traditional pairwise correlations (e.g.,

of a member of specific taxon and an ecosystem output) from classical studies.

Notably, previous *Tara* studies have revealed prokaryotic and eukaryotic DNA virus abundances to provide biological proxies for estimating carbon export (*10*, *44*), and one even identified eukaryotic virus abundances as predictive for carbon export efficiency (*44*). However, the RNA virus diversity and abundance analyses presented here represent major advances: (i) our ecological unit and abundance calculation methods (from contigs to high-quality genomes) were extensively evaluated and found to be robust and suitable for sensitive ecological analyses (figs. S4 and S5); (ii) our analyses were composed purely of RNA viruses because of capturing 25-fold more data that are not dominated by eukaryotic dsDNA viruses; and (iii) our analyses included polar waters, which are critical for carbon export (fig. S8). Together, these findings provide a roadmap for studying RNA viruses in nature, as well as evidence that RNA viruses play important roles in the ocean ecosystem.

REFERENCES AND NOTES

1. C. Costello *et al*., *Nature* **588**, 95–100 (2020).
2. P. G. Falkowski, T. Fenchel, E. F. Delong, *Science* **320**, 1034–1039 (2008).
3. C. B. Field, M. J. Behrenfeld, J. T. Randerson, P. Falkowski, *Science* **281**, 237–240 (1998).
4. P. W. Boyd, H. Claustre, M. Levy, D. A. Siegel, T. Weber, *Nature* **568**, 327–335 (2019).
5. A. Z. Worden *et al*., *Science* **347**, 1257594 (2015).
6. C. A. Suttle, *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
7. A. Culley, *Virus Res.* **244**, 84–89 (2018).
8. G. F. Steward *et al*., *ISME J.* **7**, 672–679 (2013).
9. J. A. Miranda, A. I. Culley, C. R. Schvarcz, G. F. Steward, *Environ. Microbiol.* **18**, 3714–3727 (2016).
10. L. Guidi *et al*., *Nature* **532**, 465–470 (2016).
11. M. Moniruzzaman *et al*., *Nat. Commun.* **8**, 16054 (2017).
12. Y. Tomaru *et al*., *Environ. Microbiol.* **9**, 1376–1383 (2007).
13. L. Zeigler Allen *et al*., *mSystems* **2**, e00125–e16 (2017).
14. J. A. Gustavsen, D. M. Winget, X. Tian, C. A. Suttle, *Front. Microbiol.* **5**, 703 (2014).
15. E. P. Starr, E. E. Nuccio, J. Pett-Ridge, J. F. Banfield, M. K. Firestone, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 25900–25908 (2019).
16. M. Shi *et al*., *Nature* **540**, 539–543 (2016).
17. Y. I. Wolf *et al*., *mBio* **9**, e02329–e18 (2018).
18. C.-X. Li *et al*., *eLife* **4**, e05378 (2015).
19. A. A. Zayed *et al*., *Science* **376**, 156–162 (2022).
20. A. Longhurst, S. Sathyendranath, T. Platt, C. Caverhill, *J. Plankton Res.* **17**, 1245–1271 (1995).
21. A. C. Gregory *et al*., *Cell* **177**, 1109–1123.e14 (2019).
22. G. Salazar *et al*., *Cell* **179**, 1068–1083.e21 (2019).
23. K. Bandara, Ø. Varpe, L. Wijewardene, V. Tverberg, K. Eiane, *Biol. Rev. Camb. Philos. Soc.* **96**, 1547–1589 (2021).
24. G. Sommeria-Klein *et al*., *Science* **374**, 594–599 (2021).
25. S. Chaffron *et al*., *Sci. Adv.* **7**, eabg1921 (2021).
26. M. Słowakiewicz *et al*., *Geochim. Cosmochim. Acta* **292**, 482–498 (2021).
27. M. Sadeghi, Y. Tomaru, T. Ahola, *Viruses* **13**, 362 (2021).
28. K. F. Edwards, G. F. Steward, C. R. Schvarcz, *Ecol. Lett.* **24**, 363–373 (2021).
29. S. Sunagawa *et al*., *Science* **348**, 1261359 (2015).
30. M. J. Costello, C. Chaudhary, *Curr. Biol.* **27**, R511–R527 (2017).
31. D. Righetti, M. Vogt, N. Gruber, A. Psomas, N. E. Zimmermann, *Sci. Adv.* **5**, eaau6253 (2019).
32. M. R. Willig, D. M. Kaufman, R. D. Stevens, *Annu. Rev. Ecol. Evol. Syst.* **34**, 273–309 (2003).
33. F. M. Ibarbalz *et al*., *Cell* **179**, 1084–1097.e21 (2019).
34. J. B. Emerson, B. C. Thomas, K. Andrade, K. B. Heidelberg, J. F. Banfield, *Appl. Environ. Microbiol.* **79**, 6755–6764 (2013).
35. A. C. Gregory *et al*., *Cell Host Microbe* **28**, 724–740.e8 (2020).
36. E. A. Gould, *Mol. Biotechnol.* **13**, 57–66 (1999).
37. C. J. Carlson, C. M. Zipfel, R. Garnier, S. Bansal, *Nat. Ecol. Evol.* **3**, 1070–1075 (2019).
38. M. Breitbart, L. R. Thompson, C. A. Suttle, M. B. Sullivan, *Oceanography (Wash. D.C.)* **20**, 135–139 (2007).
39. G. Meyers, T. Rümenapf, H.-J. Thiel, *Nature* **341**, 491–491 (1989).
40. A. M. Burroughs, L. Aravind, *Nucleic Acids Res.* **44**, 8525–8555 (2016).
41. A. R. Fehr, G. Jankevicius, I. Ahel, S. Perlman, *Trends Microbiol.* **26**, 598–610 (2018).
42. F. Vincent, U. Sheyn, Z. Porat, D. Schatz, A. Vardi, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2021586118 (2021).
43. T. E. G. Biggs, J. Huisman, C. P. D. Brussaard, *ISME J.* **15**, 3615–3622 (2021).
44. H. Kaneko *et al*., *iScience* **24**, 102002 (2020).
45. A. A. Zayed, J. M. Wainaina, G. Dominguez-Huerta, Cryptic and abundant marine viruses at the evolutionary origins of Earth's RNA virome. CyVerse Data Commons (2021).

ACKNOWLEDGMENTS

We thank A. Crane (Integrated Research Facility at Fort Detrick, National Institute of Allergy and Infectious Diseases, National Institutes of Health) for critically editing the manuscript. *Tara* Oceans would not exist without the leadership of the *Tara* Oceans Foundation and its sponsors, and the continuous support of 23 institutes (expeditionary support is detailed in the Supplementary Text).
Funding: The virus-specific work presented here was supported in part through the following: US National Science Foundation (awards OCE 1829831, ABI 1759874, and DBI 2022070); The Gordon and Betty Moore Foundation (award 3790); The Ohio Supercomputer and Ohio State University's Center of Microbiome Science to M.B.S.; Ramon-Areces Foundation Postdoctoral Fellowship to G.D.-H.; and France Génomique (ANR-10-INBS-09) to P.W. This work was supported in part through Laulima Government Solutions, LLC, prime contract with the US National Institute of Allergy and Infectious Diseases (NIAID) under contract no. HHSN272201800013C. J.H.K. performed this work as an employee of Tunnell Government Services (TGS), a subcontractor of Laulima Government Solutions, LLC, under contract no. HHSN272201800013C. **Author contributions:** G.D.-H., A.A.Z., and M.B.S. planned and supervised the work, interpreted the results, and wrote the manuscript with inputs from all authors. G.D.-H., A.A.Z., J.M.W., J.G., F.T., A.A.P., B.B., M.M., O.Z., E.P., E.D., and S.C. developed and/or implemented the informatic analyses. A.A., J.-M.A., Q.C., C.d.S., K.L., J.P., A.A.Z., P.W., and *Tara* Oceans coordinators all contributed to expeditionary infrastructure needed for Global Ocean sampling, sample processing, and/or previously published data resource development. C.B., D.E., E.K., and H.O. provided domain expertise on Global Ocean ecology. L.G., J.H.K., and A.C. provided domain expertise on carbon export, the ecological unit, and RNA virus ecology, respectively. All authors read and commented on the manuscript and approved it in its final form. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The authors declare that all data reported here are fully and freely available from the date of publication without restrictions and that all of the analyses, publications, and ownership of data are free from legal entanglement or restriction by the various nations whose waters were sampled during the *Tara* Oceans Expeditions. This article is contribution no. 133 of *Tara* Oceans. Processed data are publicly available through iVirus (*45*), including all metatranscriptome assemblies, RNA virus contigs, and RNA vOTUs. Scripts used to generate figures are uploaded to the MAVERICKlab bitbucket page (https://bitbucket.org/MAVERICLab/global-rna-virus-ecology-2022/). **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.abn6358
Materials and Methods
Supplementary Text
Figs. S1 to S8
Tables S1 to S7
References (*46–124*)
MDAR Reproducibility Checklist
Data S1 to S2

View/request a protocol for this paper from *Bio-protocol*.

Submitted 20 December 2021; accepted 4 May 2022
10.1126/science.abn6358

QUANTUM SIMULATION

Quantum optimization of maximum independent set using Rydberg atom arrays

S. Ebadi[1]†, A. Keesling[1,2]†, M. Cain[1]†, T. T. Wang[1], H. Levine[1]‡, D. Bluvstein[1], G. Semeghini[1], A. Omran[1,2], J.-G. Liu[1,2], R. Samajdar[1], X.-Z. Luo[2,3,4], B. Nash[5], X. Gao[1], B. Barak[5], E. Farhi[6,7], S. Sachdev[1,8], N. Gemelke[2], L. Zhou[1,9], S. Choi[7], H. Pichler[10,11], S.-T. Wang[2], M. Greiner[1]*, V. Vuletić[12]*, M. D. Lukin[1]*

Realizing quantum speedup for practically relevant, computationally hard problems is a central challenge in quantum information science. Using Rydberg atom arrays with up to 289 qubits in two spatial dimensions, we experimentally investigate quantum algorithms for solving the maximum independent set problem. We use a hardware-efficient encoding associated with Rydberg blockade, realize closed-loop optimization to test several variational algorithms, and subsequently apply them to systematically explore a class of graphs with programmable connectivity. We find that the problem hardness is controlled by the solution degeneracy and number of local minima, and we experimentally benchmark the quantum algorithm's performance against classical simulated annealing. On the hardest graphs, we observe a superlinear quantum speedup in finding exact solutions in the deep circuit regime and analyze its origins.

Combinatorial optimization is ubiquitous in many areas of science and technology. Many such problems have been shown to be computationally hard and form the basis for understanding complexity classes in modern computer science (*1*). The use of quantum machines to accelerate solving such problems has been theoretically explored for over two decades with a variety of quantum algorithms (*2*–*4*). Typically, a relevant cost function is encoded in a quantum Hamiltonian (*5*), and its low-energy state is sought starting from a generic initial state, either through an adiabatic evolution (*2*) or a variational approach (*3*), via closed optimization loops (*6*, *7*). The computational performance of such algorithms has been investigated theoretically (*4*, *8*–*13*) and experimentally (*14*–*16*) in small quantum systems with shallow quantum circuits, or in systems lacking the many-body coherence believed to be central for quantum advantage (*17*, *18*). However, these studies offer only limited insights into algorithms' performances in the most interesting regime involving large system sizes and high circuit depths (*19*, *20*).

Here we use a quantum device based on coherent, programmable arrays of neutral atoms trapped in optical tweezers to investigate quantum optimization algorithms for systems ranging from 39 to 289 qubits, and effective depths sufficient for the quantum correlations to spread across the entire graph. Specifically, we focus on maximum independent set, a paradigmatic NP-hard optimization problem (*21*). It involves finding the largest independent set of a graph—a subset of vertices such that no edges connect any pair in the set. An important class of such maximum independent set problems involves unit disk graphs, which are defined by vertices on a two-dimensional plane with edges connecting all pairs of vertices within a unit distance of one another (Fig. 1, A and B). Such instances arise naturally in problems associated with geometric constraints that are important for many practical applications, such as modeling wireless communication networks (*22*, *23*). Although there exist polynomial-time classical algorithms to find approximate solutions to the maximum independent set problem on such graphs (*24*), solving the problem exactly is known to be NP-hard in the worst case (*23*, *25*).

Maximum independent set on Rydberg atom arrays

Our approach uses a two-dimensional atom array described previously (*26*). Excitation from a ground state $|0\rangle$ into a Rydberg state $|1\rangle$ is utilized for hardware-efficient encoding of the unit disk maximum independent set problem (*27*). For a particular graph, we create a geometric configuration of atoms using optical tweezers such that each atom represents a vertex. The edges are drawn according to the unit disk criterion for a unit distance given by the Rydberg blockade radius R_b (Fig. 1C), the distance within which excitation of more than one atom to the Rydberg state is prohibited because of strong interactions (*28*). The Rydberg blockade mechanism thus restricts the evolution primarily to the subspace spanned by the states that obey the independent set constraint of the problem graph. Quantum algorithms for optimization are implemented via global atomic excitation using homogeneous laser pulses with a time-varying Rabi frequency (and a time-varying phase) $\Omega(t)e^{i\phi(t)}$ and detuning $\Delta(t)$ (Fig. 1D). The resulting quantum dynamics is governed by the Hamiltonian $H = H_q + H_{cost}$, with the quantum driver H_q and the cost function H_{cost} given by

$$H_q = \frac{h}{2}\sum_i \left[\Omega(t)e^{i\phi(t)}|0\rangle_i\langle 1| + \text{h.c.}\right],$$
$$H_{cost} = -h\Delta(t)\sum_i n_i + \sum_{i<j} V_{ij} n_i n_j \qquad (1)$$

where $n_i = |1\rangle_i\langle 1|$, and $V_{ij} = V_0/(|r_i - r_j|)^6$ is the interaction potential that sets the blockade radius R_b and determines the connectivity of the graph. For a positive laser detuning Δ, the many-body ground state of the cost function Hamiltonian maximizes the total number of qubits in the Rydberg state under the blockade constraint, corresponding to the largest independent set MIS(G) (hereafter MIS) of the underlying unit disk graph G (*27*) (Fig. 1E). Even with the finite blockade energy and long-range interaction tails, we empirically find that the ground states of H_{cost} still encode an MIS for the ensemble of graphs studied here [see (*25*, *27*)].

Variational optimization via a closed quantum-classical loop

In the experiment, we deterministically prepare graphs with vertices occupying 80% of an underlying square lattice, with the blockade extending across nearest and next-nearest (diagonal) neighbors (Fig. 1C). This allows us to explore a class of nonplanar graphs for which finding the exact solution of MIS is NP-hard for worst-case instances (*25*). To prepare quantum states with a large overlap with the MIS solution space, we use a family of variational quantum optimization algorithms using a quantum-classical optimization loop. We place atoms at positions defined by the vertices of the chosen graph, initialize them in state $|0\rangle$, and implement a coherent quantum evolution corresponding to the specific choice of variational parameters (Fig. 1D). Subsequently, we sample the wave function with a projective measurement and determine the

[1]Department of Physics, Harvard University, Cambridge, MA 02138, USA. [2]QuEra Computing Inc., Boston, MA 02135, USA. [3]Department of Physics and Astronomy, University of Waterloo, Waterloo N2L 3G1, Canada. [4]Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada. [5]School of Engineering and Applied Science, Harvard University, Cambridge, MA 02138, USA. [6]Google Quantum AI, Venice, CA 90291, USA. [7]Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [8]School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540, USA. [9]Walter Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, CA 91125, USA. [10]Institute for Theoretical Physics, University of Innsbruck, A-6020 Innsbruck, Austria. [11]Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, A-6020 Innsbruck, Austria. [12]Department of Physics and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
***Corresponding author. Email: greiner@physics.harvard.edu (M.G.); vuletic@mit.edu (V.V.); lukin@physics.harvard.edu (M.D.L.)**
†These authors contributed equally to this work. ‡Present address: AWS Center for Quantum Computing, Pasadena, CA 91125, USA.

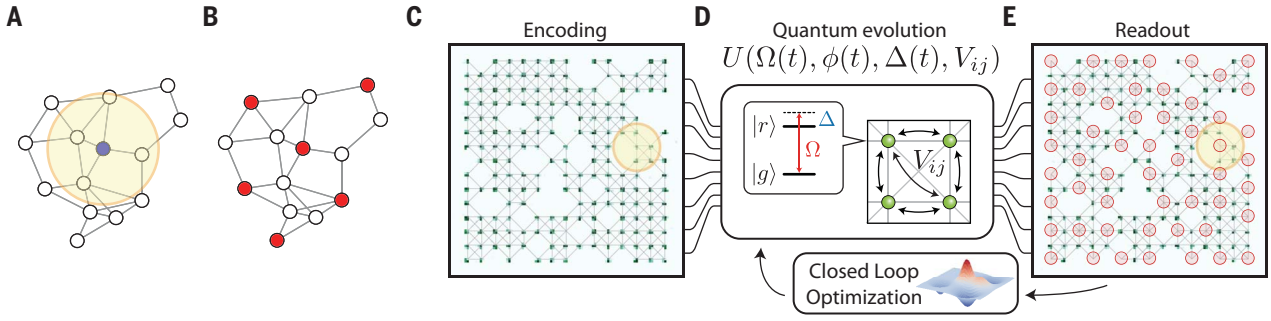


Fig. 1. Hardware-efficient encoding of the maximum independent set using Rydberg atom arrays. (**A**) An example of a unit disk graph, with any single vertex (e.g., the blue vertex) being connected to all other vertices within a disk of unit radius. (**B**) A corresponding MIS solution (denoted by the red nodes). (**C**) The maximum independent set problem is encoded with atoms placed at the vertices of the target graph and with interatomic spacing chosen such that the unit disk radius of the graph corresponds to the Rydberg blockade radius. Shown is an example fluorescence image of atoms, with gray lines added to indicate edges between connected vertices. (**D**) The system undergoes coherent quantum many-body evolution under a programmable laser drive $[\Omega(t), \phi(t), \Delta(t)]$ and long-range Rydberg interactions V_{ij}. (**E**) A site-resolved projective measurement reads out the final quantum many-body state, with atoms excited to the Rydberg state (red circles) corresponding to vertices forming an independent set. A classical optimizer uses the results to update the parameters of the quantum evolution $[\Omega(t), \phi(t), \Delta(t)]$ to maximize a figure of merit for finding an MIS.

size of the output independent set by counting the number of qubits in $|1\rangle$, using classical post-processing to remove blockade violations and reduce detection errors (*25*) (Fig. 1E). This procedure is repeated multiple times to estimate the mean independent set size $\langle\sum_i n_i\rangle$ of the sampled wave function, the approximation ratio $R \equiv \langle\sum_i n_i\rangle/|\mathrm{MIS}|$, and the probability P_{MIS} of observing an MIS (where |MIS| denotes the size of an MIS of the graph). The classical optimizer tries to maximize $\langle\sum_i n_i\rangle$ by updating the variational parameters in a closed-loop hybrid quantum-classical optimization protocol (*25*) (Fig. 1D).

We test two algorithm classes, defined by different parametrizations of the quantum driver and the cost function in Eq. 1. The first approach consists of resonant ($\Delta = 0$) laser pulses of varying durations τ_i and phases ϕ_i (Fig. 2A). This algorithm closely resembles the canonical quantum approximate optimization algorithm (QAOA) (*3*), but instead of exact single-qubit rotations, resonant driving generates an effective many-body evolution within the subspace of independent sets associated with the blockade constraint (*25*). Phase jumps between consecutive pulses implement a global phase gate (*29*), with a phase shift proportional to the cost function of the maximum independent set problem in the subspace of independent sets (see eq. S2). Taken together, these implement the QAOA, where each pulse duration τ_i and phase ϕ_i are used as a variational parameters.

The performance of QAOA as a function of depth p (the number of pulses) is shown in Fig. 2B for an instance of a 179-vertex graph embedded in a 15 × 15 lattice. We find that the approximation ratio grows as a function of the number of pulses up to $p = 4$, and increasing the depth further does not appear to lead to better performance (Fig. 2B). As discussed in (*25*), we attribute these performance limitations to the difficulty of finding the optimal QAOA parameters for large depths within a limited number of queries to the experiment, leakage out of the independent set subspace during resonant excitation due to imperfect blockade associated with the finite interaction energy between next-nearest neighbors, and laser pulse imperfections.

The second approach is a variational quantum adiabatic algorithm (VQAA) (*2*, *30*), related to methods previously used to prepare quantum many-body ground states (*26*, *31*, *32*). In this approach, we sweep the detuning Δ from an initial negative detuning Δ_0 to a final large positive value Δ_f at constant Rabi frequency Ω, along a piecewise-linear schedule characterized by a total number of segments f, the duration τ_i of each, and the end detuning Δ_i of each segment. Moreover, we turn on the coupling Ω in duration τ_Ω and smoothen the detuning sweep using a low-pass filter with a characteristic filter time τ_Δ (Fig. 2C), both of which minimize nonadiabatic excitations and serve as additional variational parameters. For this evolution, we define an effective circuit depth $\tilde{p}$ as the duration of the sweep ($T = \tau_1 + \ldots + \tau_f$) in units of the π-pulse time τ_π, which is the time required to perform a spin flip operation.

We find that with only three segments optimized for an effective depth of $\tilde{p} = 10$ (Fig. 2D inset), the optimizer converges to a pulse that substantially outperforms the QAOA approach described above. Furthermore, the optimized pulse shows a better performance compared to a linear (one-segment) detuning sweep of the same $\tilde{p}$ (Fig. 2D). We find that similar pulse shapes produce high approximation ratios for a variety of graphs (see, e.g., fig. S8C), consistent with theoretical predictions of pulse shape concentration (*20*, *25*, *33*, *34*). At large sweep times ($\tilde{p} > 15$), we observe a turn-around in the performance likely associated with decoherence (*25*). For the remainder of this work, we focus on the quantum adiabatic algorithm for solving maximum independent set.

Quantum optimization on different graphs

The experimentally optimized quasi-adiabatic sweep (depicted in Fig. 2D) was applied to 115 randomly generated graphs of various sizes ($N = 80$ to 289 vertices). For graphs of the same size ($N = 180$), the approximation error $1 - R$ decreases and the probability of finding an MIS solution P_{MIS} increases with the effective circuit depth at early times, with the former showing a scaling consistent with a power-law relation for short effective depths (Fig. 3A and fig. S15) (*25*). We find a strong correlation between the performance of the quantum algorithm on a given graph and its total number of MIS solutions, which we refer to as the MIS degeneracy $D_{|\mathrm{MIS}|}(G)$ (hereafter $D_{|\mathrm{MIS}|}$). This quantity is calculated classically using a tensor network algorithm (*25*, *35*) and varies by nine orders of magnitude across different 180-vertex graphs. We observe a clear logarithmic relation between $D_{|\mathrm{MIS}|}$ and the approximation error $1 - R$, accompanied by a nearly three-orders-of-magnitude variation of P_{MIS} at a fixed depth $\tilde{p} = 20$ (Fig. 2B). P_{MIS} does not scale linearly with the MIS degeneracy, as would be the case for a naive algorithm that samples solutions at random. Figure 2C shows the sharp collapse of $1 - R$ as a function of the logarithm of the MIS degeneracy normalized by the graph size, $\rho \equiv \log(D_{|\mathrm{MIS}|})/N$. This quantity, a measure of MIS degeneracy density, determines the hardness in approximating solutions for the quantum algorithm at shallow depths.

These observations can be modeled as resulting from a Kibble-Zurek–type mechanism

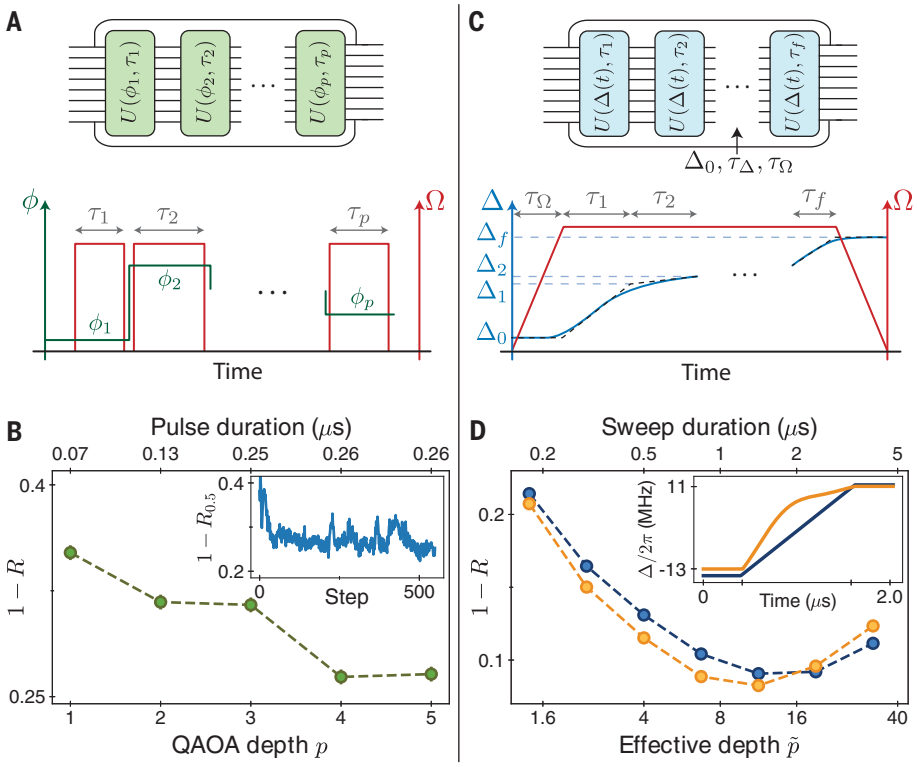


Fig. 2. Testing variational quantum algorithms. (**A**) Implementation of the quantum approximate optimization algorithm (QAOA), consisting of sequential layers of resonant pulses with variable duration τ_i and laser phase ϕ_i. (**B**) Variational optimization of QAOA parameters results in a decrease in approximation error $1 - R$, up to depth $p = 4$ (inset: example performance of quantum-classical closed-loop optimization at $p = 5$). Approximation error calculated using the top 50 percentiles of independent set sizes ($1 - R_{0.5}$) is used as the figure of merit to reduce effects of experimental imperfections on the optimization procedure (*25*). (**C**) Quantum evolution can also be parametrized as a variational quantum adiabatic algorithm (VQAA) using a quasi-adiabatic pulse with a piecewise-linear sweep of detuning $\Delta(t)$ at constant Rabi coupling $\Omega(t)$. $\Omega(t)$ is turned on and off within τ_Ω, and a low-pass filter with time scale τ_Δ is used to smoothen the $\Delta(t)$ sweep. (**D**) Performance of a rescaled piecewise-linear sweep as a function of its effective depth $\tilde{p} = (\tau_1 + \ldots + \tau_f)/\tau_\pi$. Variational optimization of a three-segment (orange) piecewise-linear pulse (optimized for $\tilde{p} = 10$) improves on the performance of a simple one-segment linear (blue) pulse, as well as the best results from QAOA (inset: detuning sweep profiles for one-segment (blue) and three-segment (orange) optimized pulses, for a total pulse duration of 2.0 μs). Error bars for approximation ratio R are the SEM here and throughout the text and are smaller than the points.

where the quantum algorithm locally solves the graph in domains whose sizes are determined by the evolution time and speed at which quantum information propagates (*36*, *37*). We show that the scaling of the approximation error with depth can originate from the conflicts between local solutions at the boundaries of these independent domains (*25*). In graphs with a large degeneracy density ρ, there may exist many MIS configurations that are compatible with the local ordering in these domains. This provides a possible mechanism to reduce domain walls at their boundaries (fig. S14) and decrease the approximation error. Such a scenario would predict a linear relation between $1 - R$ and ρ at a fixed depth, which is consistent with our observations (Fig. 2C and fig. S15).

Benchmarking against simulated annealing

To benchmark the results of the quantum optimization against a classical algorithm, we use simulated annealing (SA) (*38*). It seeks to minimize the energy of a cost Hamiltonian by thermally cooling a system of classical spins while maintaining thermal equilibrium. Although some specifically tailored state-of-the-art algorithms (*24*, *39*) may have better performance than SA in solving the maximum independent set problem, we have chosen SA for extensive benchmarking because similar to the quantum algorithms used, it is a general-purpose algorithm that only relies on information from the cost Hamiltonian for solving the problem. Our highly optimized variant of SA stochastically updates local clusters of spins using the Metropolis-Hastings (*40*) update rule, rejecting energetically unfavorable updates with a probability dependent on the energy cost and the instantaneous temperature (*25*). We use collective updates under the MIS Hamiltonian cost function (eq. S15), which applies an optimized uniform interaction energy to each edge, penalizing states that violate the independent set criterion (*25*). The annealing depth p_{SA} is defined as the average number of attempted updates per spin.

We compare the quantum algorithm and SA on two metrics: the approximation error $1 - R$, and the probability of sampling an exact solution P_{MIS}, which determines the inverse of time-to-solution. As shown in Fig. 4A, for relatively shallow depths and moderately hard graphs, optimized SA results in approximation errors similar to those observed on the quantum device. In particular, we find that the hardness in approximating the solution for short SA depths is also controlled by degeneracy density ρ (fig. S18, A and B). However, some graph instances appear to be considerably harder for SA compared to the quantum algorithm at higher depths (see, e.g., gold and purple curves in Fig. 4A).

Detailed analysis of the SA dynamics for graphs with low degeneracy densities ρ reveals that for some instances, the approximation ratio displays a plateau at $R = (|\text{MIS}| - 1)/|\text{MIS}|$, corresponding to independent sets with one less vertex than an MIS (Fig. 4A, gold and purple solid lines). Graphs displaying this behavior have a large number of local minima with independent set size $|\text{MIS}| - 1$, in which SA can be trapped up to large depths. By analyzing the dynamics of SA at low temperatures as a random walk among $|\text{MIS}| - 1$ and $|\text{MIS}|$ configurations (Fig. 4D), we show in (*25*) that the ability of SA to find a global optimum is limited by the ratio of the number of suboptimal independent sets of size $|\text{MIS}| - 1$ to the number of ways to reach global minima, resulting in a "hardness parameter" $\mathcal{HP} = D_{|\text{MIS}|-1}/(|\text{MIS}|D_{|\text{MIS}|})$ (Fig. 4E). This parameter lower bounds the mixing time for the Markov chain describing the SA dynamics at low temperatures (eq. S19), and it appears to increase exponentially with the square root of the system size for the hardest graphs (fig. S11). This suggests that a large number of local minima cause SA to take an exponentially long time to find an MIS for the hardest cases as N grows. If SA performance saturates this lower bound, consistent with numerics (fig. S19), its runtime to find an MIS is polynomially related to the best known exact classical algorithms (*41*).

Quantum speedup on the hardest graphs

We now turn to study the algorithms' ability to find exact solutions on the hardest graphs (with up to $N = 80$), chosen from graphs in the top two percentile of the hardness parameter $\mathcal{HP}$ (fig. S11). We find that for some

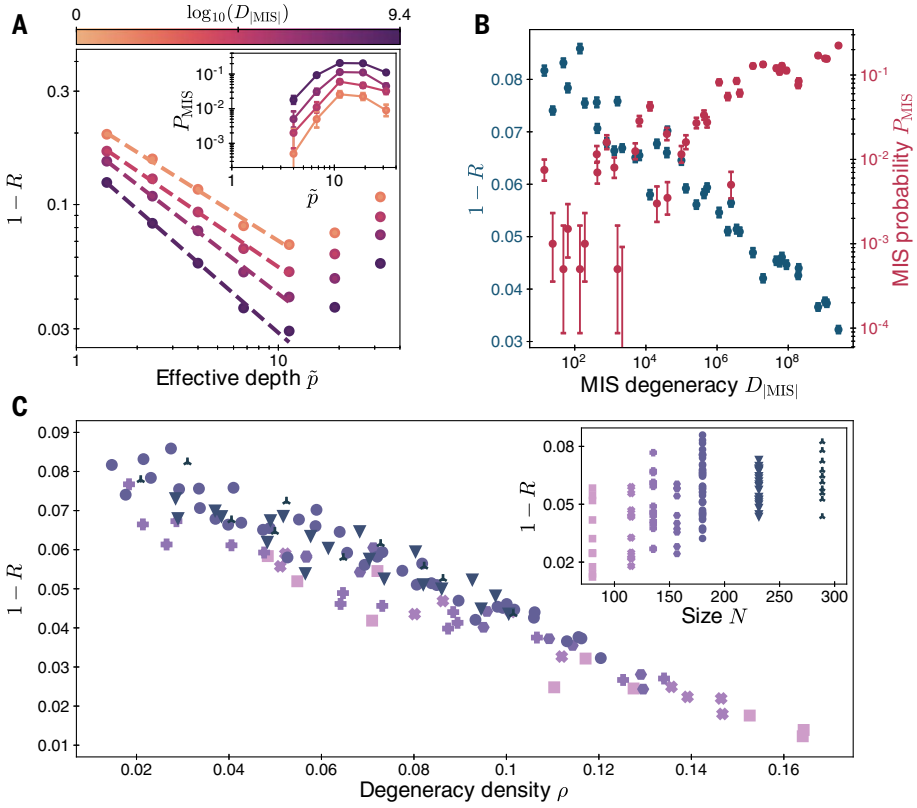


Fig. 3. Quantum algorithm performance across different graphs. (**A**) The approximation error $1 - R$ for an optimized quasi-adiabatic sweep plotted as a function of effective depth $\tilde{p}$ on four graphs of the same size (N = 180 vertices), showing strong dependence on the number of MIS solutions (MIS degeneracy) $D_{|\mathrm{MIS}|}$ (inset: corresponding MIS probability P_{MIS} versus $\tilde{p}$). (**B**) At a fixed depth $\tilde{p}$ = 20, $1 - R$ and P_{MIS} for various 180-vertex graphs are strongly correlated with $D_{|\mathrm{MIS}|}$. (**C**) At the same effective depth $\tilde{p}$ = 20, $1 - R$ for 115 graphs of different sizes (N = 80 to 289) and MIS degeneracies $D_{|\mathrm{MIS}|}$ exhibit universal scaling with the degeneracy density $\rho \equiv \log(D_{|\mathrm{MIS}|})/N$ (inset: data plotted as a function of N). Error bars for P_{MIS}, here and throughout the text, denote the 68% confidence interval.

of these graphs (e.g., gold curves in Fig. 4, A to C), the quantum algorithm quickly approaches the correct solutions, reducing the average Hamming distance (number of spin flips normalized by N) to the closest MIS and increasing P_{MIS}, while SA remains trapped in local minima at a large Hamming distance from any MIS. For other instances (e.g., purple curves in Fig. 4, A to C), both the quantum algorithm and SA have difficulty finding the correct solution. Moreover, in contrast to our earlier observations suggesting variational parameter concentration for generic graphs, we find that for these hard instances, the quantum algorithm needs to be optimized for each graph individually by scanning the slowdown point of the detuning sweep $\Delta(t)$ to maximize P_{MIS} (Fig. 5, A and B, and fig. S9) (*25*).

Figure 4E shows the resulting highest P_{MIS} reached within a depth of 32 for each hard graph instance as a function of the classical hardness parameter $\mathcal{HP}$. For simulated annealing, we find the scaling $P_{\mathrm{MIS}} = 1 - \exp(-C\mathcal{HP}^{-1.03(4)})$, where C is a positive fitted constant, which is in good agreement with theoretical expectations (*25*). Although for many instances the quantum algorithm outperforms SA, there are significant instance-by-instance variations, and on average, we observe a similar scaling $P_{\mathrm{MIS}} = 1 - \exp(-C\mathcal{HP}^{-0.95(15)})$ (dashed red line).

To understand these observations, we carried out detailed analyses of both classical and quantum algorithms' performance for hard graph instances. Specifically, in (*25*) we show that for a broad class of SA algorithms with both single-vertex and correlated updates, the scaling is at best $P_{\mathrm{MIS}} = 1 - \exp(-C\mathcal{HP}^{-1})$ (where C generally could have polynomial dependence on the system size), indicating that the observed scaling of our version of SA is close to optimal. To gain insight into the origin of the quantum scaling, we numerically compute the minimum energy gap $\delta_{\min}$ during the adiabatic evolution using density-matrix renormalization group (Fig. 5A) (*25*). Figure 5C shows that the performance of the quantum algorithm is mostly well described by quasi-adiabatic evolution with transition probability out of the ground state governed by the minimum energy gap, according to the Landau-Zener formula $P_{\mathrm{MIS}} = 1 - \exp\left(-A\delta_{\min}^{\eta}\right)$ for a constant A, and η = 1.2(2) (*42*). This observation suggests that our quantum algorithm achieves near-maximum efficiency, consistent with the smallest possible value of $\eta = 1$ obtained for optimized adiabatic following (*43*).

By focusing only on instances with large enough spectral gaps such that the evolution time T obeys the "speed limit" determined by the uncertainty principle ($\delta_{\min} > 1/T$) associated with Landau-Zener scaling (*42*), we find an improved quantum algorithm scaling $P_{\mathrm{MIS}} = 1 - \exp(-C\mathcal{HP}^{-0.63(13)})$ (Fig. 4E, solid red line). Because $1/[-\log(1 - P_{\mathrm{MIS}})] \approx 1/P_{\mathrm{MIS}}$ is proportional to the runtime sufficient to find a solution by repeating the experiment, the smaller exponent observed in the scaling for quantum algorithm ($\sim\mathcal{HP}^{1.03(4)}$ for SA and $\sim\mathcal{HP}^{0.63(13)}$ for the quantum algorithm) suggests a superlinear [with a ratio in scaling of 1.6(3)] speedup in the runtime to find an MIS, for graphs where the deep-circuit-regime ($T > 1/\delta_{\min}$) is reached. Moreover, the observed scaling is not altered by the postprocessing used on the experimental data (*25*). We emphasize that achieving this speedup requires an effective depth large enough to probe the lowest-energy many-body states of the system; by contrast, no speedup is observed for graph instances where this depth condition is not fulfilled.

Discussion and outlook

Several mechanisms for quantum speedup in combinatorial optimization problems have been previously proposed. Grover-type algorithms are known to have a quadratic speedup in comparison to brute-force classical search over all possible solutions (*44*, *45*). A quadratic quantum speedup has also been suggested for quantized SA based on discrete quantum walks (*46*, *47*). However, these methods use

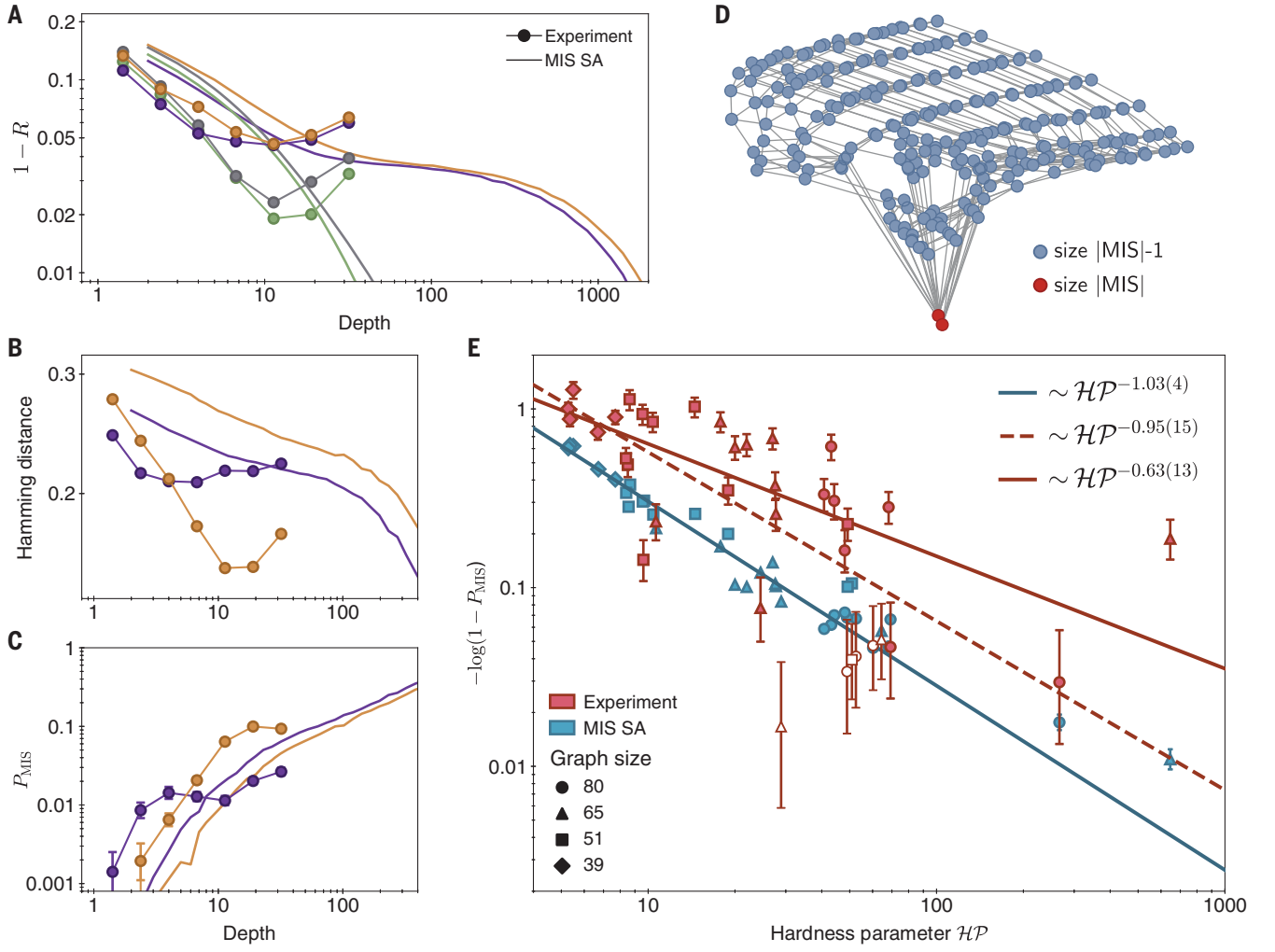


Fig. 4. Benchmarking the quantum algorithm against classical simulated annealing. (**A**) Performance of the quantum algorithm, and the optimized simulated annealing with the MIS Hamiltonian, shown as a function of depth ($\tilde{p}$ for quantum algorithm and p_{SA} for simulated annealing) for four 80-vertex graphs. Green ($\mathcal{HP}$ = 1.8, ρ = 0.13) and gray ($\mathcal{HP}$ = 2.1, ρ = 0.11) graphs are easy for the quantum and classical algorithm; however, purple ($\mathcal{HP}$ = 69, ρ = 0.08) and gold ($\mathcal{HP}$ = 68, ρ = 0.06 are significantly harder and show a plateau at R = (|MIS| − 1)/|MIS|, i.e., independent sets with one less vertex than an MIS. (**B** and **C**) One of the hard graphs (gold) shows much better quantum scaling of average normalized Hamming distance to the closest MIS, and MIS probability (P_{MIS}) compared to the other graph (purple). By contrast, the performance of SA (lines) remains similar between the two graphs. (**D**) Configuration graph of independent sets of size |MIS| and |MIS| − 1 for an example 39-vertex graph ($\mathcal{HP}$ = 5), where the edges connect two configurations if they are separated by one step of simulated annealing. At low temperatures, simulated annealing finds an MIS solution by a random walk on this configuration graph. (**E**) −log(1 − P_{MIS}) for instance-by-instance optimized quantum algorithm (crimson) and simulated annealing (teal) reached within a depth of 32, for 36 graphs selected from the top two percentile of hardness parameter $\mathcal{HP}$ for each size. Power-law fits to the SA (teal, $\sim\mathcal{HP}^{-1.03(4)}$) and the quantum data (dashed crimson line, $\sim\mathcal{HP}^{-0.95(15)}$) are used to compare scaling performance with graph hardness $\mathcal{HP}$. The error in the power-law exponents from the fit is the combination of statistical errors and the error in the least-squares fit. If only graphs with minimum energy gaps large enough to be resolved in the duration of the quantum evolution are considered ($\delta_{min} > 1/T$, excluding hollow data points), the fit (solid crimson line) shows a superlinear speedup $\sim\mathcal{HP}^{-0.63(13)}$ over optimized simulated annealing.

specifically constructed circuits and are not directly applicable to the algorithms implemented here. In addition, the following mechanisms can contribute to the speedup observed in our system. The quantum algorithm's performance in the observed regime appears to be mostly governed by the minimum energy gap δ_{min} (Fig. 5C). We show that under certain conditions, one can achieve coherent quantum enhancement for the minimum gap resulting in a quadratic speedup via $\delta_{min} \sim \mathcal{HP}^{-1/2}$ (*25*). In practice, however, we find that the minimum energy gap does not always correlate with the classical hardness parameter $\mathcal{HP}$, as is evident in the spread of the quantum data in Fig. 4E (see also fig. S21). Some insights into these effects can be gained by a more direct comparison of the quantum algorithm with SA using the same cost function corresponding to the Rydberg Hamiltonian (*25*) (Fig. 5D). Although the observed power-law scaling supports the possibility of a nearly quadratic speedup for instances in the deep circuit regime ($\delta_{min} > 1/T$), it is an open question whether such a speedup can be extended, with a guarantee, in all instances. Finally, it is possible that δ_{min} alone does not fully determine the quantum performance, as suggested by the data points that deviate from the Landau-Zener prediction in Fig. 5C, where

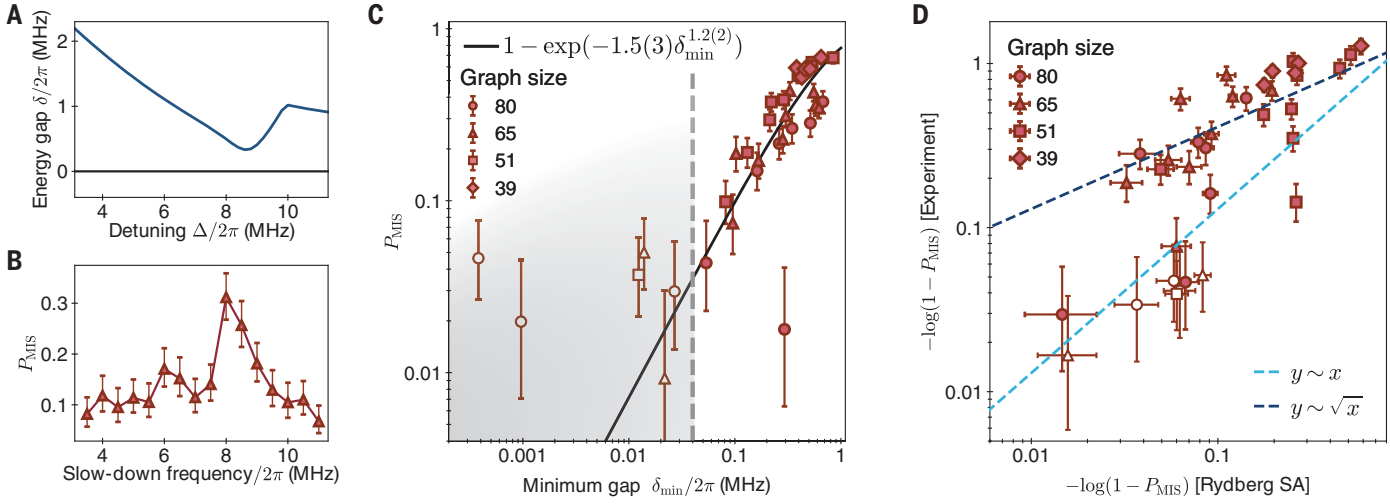


Fig. 5. Understanding hardness for the quantum algorithm. (**A**) Energy gap between the ground (black) and first-excited (blue) states, calculated using the density matrix renormalization group (DMRG) for a graph of 65 atoms. (**B**) To maximize P_{MIS} for hard graphs, the frequency at which the detuning sweep is slowed down is varied (see fig. S9). The largest P_{MIS} corresponds to a slow-down frequency close to the location of the minimum gap. (**C**) Measured P_{MIS} for a fixed effective depth $\tilde{p}$ = 32 as a function of the calculated minimum gap δ_{min}. For many instances, the relation is well described by the Landau-Zener prediction for quasi-adiabatic ground state preparation. The shaded region corresponds to when the gap is too small ($\delta_{\mathrm{min}} < 1/T$) to be properly resolved relative to the quantum evolution time, and points in this region are excluded from the fit both here and in the solid crimson line in Fig. 4E. (**D**) Scaling of $-\log(1 - P_{\mathrm{MIS}})$ observed in the experiment versus in simulated annealing under the classical Rydberg cost function, eq. S14, for best P_{MIS} reached within a depth of 32. These results are consistent with a nearly quadratic speedup for a subset of graphs where $\delta_{\mathrm{min}} > 1/T$.

enhancement through diabatic effects could be possible (*34*, *48*).

Although the scaling speedup observed here suggests a possibility of quantum advantage in runtime, to achieve practical runtime speedups over specialized state-of-the-art heuristic algorithms [e.g., (*39*)], qubit coherence, system size, and the classical optimizer loop need to be improved. The useful depth accessible via quantum evolution is limited by Rydberg-state lifetime and intermediate-state laser scattering, which can be suppressed by increasing the control laser intensity and intermediate-state detuning. Advanced error mitigation techniques such as STIRAP (*49*), as well as error correction methods, should also be explored to enable large-scale implementations. The classical optimization loop can be improved by speeding up the experimental cycle time and by using more advanced classical optimizers. Larger atom arrays can be realized by using improvements in vacuum-limited trap lifetimes and sorting fidelity.

Our results demonstrate the potential of quantum systems for the discovery of new algorithms and highlight a number of new scientific directions. It would be interesting to investigate whether instances with large Hamming distance between the local and global optima of independent set sizes |MIS| – 1 and |MIS| can be related to the overlap gap property of the solution space, which is associated with classical optimization hardness (*50*). In particular, our method can be applied to the optimization of "planted graphs," designed to maximize the Hamming distance between optimal and suboptimal solutions, which can provably limit the performance of local classical algorithms (*51*). Our approach can also be extended to beyond unit disk graphs by using ancillary atoms, hyperfine qubit encoding, and a reconfigurable architecture based on coherent transport of entangled atoms (*52*). Furthermore, local qubit addressing during the evolution can be used to both extend the range of optimization parameters and the types of optimization problems (*5*). Further analysis could elucidate the origins of classical and quantum hardness, for example, by using graph neural network approaches (*53*). Finally, similar approaches can be used to explore realizations of other classes of quantum algorithm [see, e.g., (*54*)], enabling a broader range of potential applications.

REFERENCES AND NOTES

1. M. Sipser, *Introduction to the Theory of Computation* (Course Technology, Boston, ed. 3, 2013).
2. E. Farhi, J. Goldstone, S. Gutmann, M. Sipser, Quantum Computation by Adiabatic Evolution. arXiv:quant-ph/0001106 (2000).
3. E. Farhi, J. Goldstone, S. Gutmann, A Quantum Approximate Optimization Algorithm. arXiv:1411.4028 [quant-ph] (2014).
4. T. Albash, D. A. Lidar, *Rev. Mod. Phys.* **90**, 015002 (2018).
5. A. Lucas, *Front. Phys.* **2**, 5 (2014).
6. D. Wecker, M. B. Hastings, M. Troyer, *Phys. Rev. A* **94**, 022309 (2016).
7. C. Kokail *et al.*, *Nature* **569**, 355–360 (2019).
8. F. Barahona, *J. Phys. Math. Gen.* **15**, 3241–3253 (1982).
9. V. Bapst, L. Foini, F. Krzakala, G. Semerjian, F. Zamponi, *Phys. Rep.* **523**, 127–205 (2013).
10. E. Farhi *et al.*, *Science* **292**, 472–475 (2001).
11. E. Farhi *et al.*, *Phys. Rev. A* **86**, 052334 (2012).
12. S. Knysh, *Nat. Commun.* **7**, 12370 (2016).
13. A. P. Young, S. Knysh, V. N. Smelyanskiy, *Phys. Rev. Lett.* **104**, 020502 (2010).
14. M. P. Harrigan *et al.*, *Nat. Phys.* **17**, 332–336 (2021).
15. G. Pagano *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 25396–25401 (2020).
16. T. M. Graham *et al.*, Demonstration of multi-qubit entanglement and algorithms on a programmable neutral atom quantum computer. arXiv:2112.14589 [quant-ph] (2022).
17. T. F. Rønnow *et al.*, *Science* **345**, 420–424 (2014).
18. H. G. Katzgraber, F. Hamze, Z. Zhu, A. J. Ochoa, H. Munoz-Bauza, *Phys. Rev. X* **5**, 031026 (2015).
19. E. Farhi, D. Gamarnik, S. Gutmann, The Quantum Approximate Optimization Algorithm Needs to See the Whole Graph: A Typical Case. arXiv:2004.09002 [quant-ph] (2020).
20. C.-N. Chou, P. J. Love, J. S. Sandhu, J. Shi, Limitations of Local Quantum Algorithms on Random Max-k-XOR and Beyond. arXiv:2108.06049 [quant-ph] (2021).
21. M. R. Garey, D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness* (Freeman, 1979).
22. J. Wurtz, P. Lopes, N. Gemelke, A. Keesling, S. Wang, arXiv:2205.08500 [quant-ph] (2022).
23. B. N. Clark, C. J. Colbourn, D. S. Johnson, *Discrete Math.* **86**, 165–177 (1990).
24. E. J. van Leeuwen, in *Graph-Theoretic Concepts in Computer Science*, D. Kratsch, Ed. (Springer, 2005), pp. 351–361.
25. See supplementary materials.
26. S. Ebadi *et al.*, *Nature* **595**, 227–232 (2021).
27. H. Pichler, S.-T. Wang, L. Zhou, S. Choi, M. D. Lukin, Quantum Optimization for Maximum Independent Set Using Rydberg Atom Arrays. arXiv:1808.10816 [quant-ph] (2018).
28. M. D. Lukin *et al.*, *Phys. Rev. Lett.* **87**, 037901 (2001).
29. D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, J. M. Gambetta, *Phys. Rev. A* **96**, 022330 (2017).
30. B. F. Schiffer, J. Tura, J. I. Cirac, Adiabatic Spectroscopy and a Variational Quantum Adiabatic Algorithm. arXiv:2103.01226 [quant-ph] (2021).
31. G. Semeghini *et al.*, *Science* **374**, 1242–1247 (2021).
32. P. Scholl *et al.*, *Nature* **595**, 233–238 (2021).
33. F. G. S. L. Brandao, M. Broughton, E. Farhi, S. Gutmann, H. Neven, For Fixed Control Parameters the Quantum

Approximate Optimization Algorithm's Objective Function Value Concentrates for Typical Instances. arXiv:1812.04170 [quant-ph] (2018).
34. L. Zhou, S.-T. Wang, S. Choi, H. Pichler, M. D. Lukin, *Phys. Rev. X* **10**, 021067 (2020).
35. J.-G. Liu, X. Gao, M. Cain, M. D. Lukin, S.-T. Wang, arxiv:2205.03718 [cond-mat-stat-mech] (2022).
36. W. H. Zurek, U. Dorner, P. Zoller, *Phys. Rev. Lett.* **95**, 105701 (2005).
37. E. H. Lieb, D. W. Robinson, *Commun. Math. Phys.* **28**, 251 (1972).
38. S. Kirkpatrick, C. D. Gelatt Jr., M. P. Vecchi, *Science* **220**, 671–680 (1983).
39. S. Lamm, P. Sanders, C. Schulz, D. Strash, R. F. Werneck, *J. Heuristics* **23**, 207–229 (2017).
40. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, *J. Chem. Phys.* **21**, 1087–1092 (1953).
41. F. V. Fomin, D. Kratsch, *Exact Exponential Algorithms* (Springer, ed. 1, 2010).
42. L. D. Landau, E. M. Lifshitz, *Quantum Mechanics Non-Relativistic Theory* (Pergamon, ed. 3, 1977).
43. J. Roland, N. J. Cerf, *Phys. Rev. A* **65**, 042308 (2002).
44. L. K. Grover, A fast quantum mechanical algorithm for database search. In *Proceedings of the 28th Annual ACM Symposium on Theory of Computing*, Philadelphia, 1996.
45. C. Durr, P. Hoyer, A Quantum Algorithm for Finding the Minimum. arXiv quant-ph/9607014 (1999).
46. M. Szegedy, in *45th Annual IEEE Symposium on Foundations of Computer Science* (2004), pp. 32–41.
47. R. D. Somma, S. Boixo, H. Barnum, E. Knill, *Phys. Rev. Lett.* **101**, 130504 (2008).
48. E. Crosson, E. Farhi, C. Y.-Y. Lin, H.-H. Lin, P. Shor, Different Strategies for Optimization Using the Quantum Adiabatic Algorithm. arXiv:1401.7320 [quant-ph] (2014).
49. M. Fleischhauer, A. Imamoglu, J. P. Marangos, *Rev. Mod. Phys.* **77**, 633–673 (2005).
50. D. Gamarnik, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2108492118 (2021).
51. D. Gamarnik, I. Zadik, The Landscape of the Planted Clique Problem: Dense subgraphs and the Overlap Gap Property. arXiv:1904.07174 [math.ST] (2019).
52. D. Bluvstein *et al*., *Nature* **604**, 451–456 (2022).
53. A. Sohrabizadeh, Y. Bai, Y. Sun, J. Cong, Enabling Automated FPGA Accelerator Optimization Using Graph Neural Networks. arXiv:2111.08848 [cs.ARs] (2021).
54. D. S. Wild, D. Sels, H. Pichler, C. Zanoci, M. D. Lukin, *Phys. Rev. Lett.* **127**, 100504 (2021).
55. M. Fishman, S. R. White, E. M. Stoudenmire, The ITensor Software Library for Tensor Network Calculations. arXiv:2007.14822 [cs.MS] (2020).
56. S. Ebadi, Quantum Optimization of Maximum Independent Set using Rydberg Atom Arrays, Zenodo (2022); https://doi.org/10.5281/zenodo.6462687.

ACKNOWLEDGMENTS

We thank I. Cirac, J. Cong, S. Evered, M. Kalinowski, M. Lin, T. Manovitz, M. Murphy, B. Schiffer, J. Singh, A. Sohrabizadeh, J. Tura, and D. Wild for illuminating discussions and feedback on the manuscript. **Funding:** We acknowledge financial support from the DARPA ONISQ program (grant no. W911NF2010021), the Center for Ultracold Atoms, the National Science Foundation, the Vannevar Bush Faculty Fellowship, the US Department of Energy [DE-SC0021013 and DOE Quantum Systems Accelerator Center (contract no. 7568717)], the Army Research Office MURI, QuEra Computing, and Amazon Web Services. M.C. acknowledges support from DOE CSG award fellowship (DE-SC0020347). H.L. acknowledges support from the National Defense Science and Engineering Graduate (NDSEG) fellowship. D.B. acknowledges support from the NSF Graduate Research Fellowship Program (grant DGE1745303) and The Fannie and John Hertz Foundation. G.S. acknowledges support from a fellowship from the Max Planck/Harvard Research Center for Quantum Optics. R.S. and S.S. were supported by the U.S. Department of Energy under grant DE-SC0019030. X.G. acknowledges support from a fellowship from the Max Planck/Harvard Research Center for Quantum Optics. B.B. acknowledges support from DARPA grant W911NF2010021, a Simons investigator fellowships, NSF grants CCF 1565264 and DMS-2134157, and DOE grant DE-SC0022199. H.P. acknowledges support by the Army Research Office (grant no. W911NF-21-1-0367). The DMRG calculations in this paper were performed using the ITensor package (*54*) and were run on the FASRC Odyssey cluster supported by the FAS Division of Science Research Computing Group at Harvard University. **Author contributions:** S.E., A.K., M.C., T.T.W., H.L., D.B., G.S., A.O., and J.-G. L. contributed to building the experimental setup, performing the measurements, and data analysis. M.C., J.-G. L., R. S., X.-Z. L., B. N., X. G., L. Z., S. C., H. P., and S.-T. W. contributed to theoretical analysis and interpretation. B. B., E. F., S. S., and N. G. contributed to interpretation of the observations and benchmarking studies. All authors discussed the results and contributed to the manuscript. All work was supervised by M.G., V.V., and M.D.L. **Competing interests:** N.G., M.G., V.V., and M.D.L. are cofounders and shareholders of QuEra Computing. A.K. is a shareholder and an executive at QuEra Computing. A.O. and S.-T.W. are shareholders of QuEra Computing. Some of the techniques and methods used in this work are included in pending patent applications filed by Harvard University (Patent application nos. PCT/US2018/042080 and PCT/US2019/049115). **Data and materials availability:** Code for classical tensor network algorithms for graph characterization is available at (*35*). Data and code are available on Zenodo (*56*). **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. www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo6587
Materials and Methods
Figs. S1 to S22
Table S1
References (*57*–*87*)

Submitted 17 February 2022; accepted 22 April 2022
Published online 5 May 2022
10.1126/science.abo6587

REPORT

COMMUNITY ECOLOGY

Predator control of marine communities increases with temperature across 115 degrees of latitude

Gail V. Ashton[1]*, Amy L. Freestone[1,2,3], J. Emmett Duffy[4], Mark E. Torchin[3,5], Brent J. Sewall[2], Brianna Tracy[1,6], Mariano Albano[7], Andrew H. Altieri[3,8], Luciana Altvater[9], Rolando Bastida-Zavala[10], Alejandro Bortolus[11], Antonio Brante[12,13], Viviana Bravo[3], Norah Brown[14,15], Alejandro H. Buschmann[16], Edward Buskey[17], Rosita Calderón Barrera[18], Brian Cheng[19], Rachel Collin[3], Ricardo Coutinho[9], Luis De Gracia[12,13], Gustavo M. Dias[20], Claudio DiBacco[21], Augusto A. V. Flores[22], Maria Angélica Haddad[23], Zvi Hoffman[24], Bruno Ibañez Erquiaga[25], Dean Janiak[26], Analí Jiménez Campeán[27,28], Inti Keith[18], Jean-Charles Leclerc[29,30], Orlando Pedro Lecompte-Pérez[31], Guilherme Ortigara Longo[32], Helena Matthews-Cascon[33], Cynthia H. McKenzie[34], Jessica Miller[35], Martín Munizaga[36,37,38], Lais P. D. Naval-Xavier[9], Sergio A. Navarrete[39], Carlos Otálora[31], Lilian A. Palomino-Alvarez[40,41], Maria Gabriela Palomo[42], Chris Patrick[43], Cormack Pegau[44], Sandra V. Pereda[16], Rosana M. Rocha[23], Carlos Rumbold[45], Carlos Sánchez[24], Adolfo Sanjuan-Muñoz[31], Carmen Schlöder[3], Evangelina Schwindt[46], Janina Seemann[3,47], Alan Shanks[48], Nuno Simoes[41,49,50], Luis Skinner[51], Nancy Yolimar Suárez-Mozo[40,41], Martin Thiel[36,37,38], Nelson Valdivia[30,52], Ximena Velez-Zuazo[53], Edson A. Vieira[32], Bruno Vildoso[54], Ingo S. Wehrtmann[55], Matt Whalen[4,14,56], Lynn Wilbur[57], Gregory M. Ruiz[1]*

Early naturalists suggested that predation intensity increases toward the tropics, affecting fundamental ecological and evolutionary processes by latitude, but empirical support is still limited. Several studies have measured consumption rates across latitude at large scales, with variable results. Moreover, how predation affects prey community composition at such geographic scales remains unknown. Using standardized experiments that spanned 115° of latitude, at 36 nearshore sites along both coasts of the Americas, we found that marine predators have both higher consumption rates and consistently stronger impacts on biomass and species composition of marine invertebrate communities in warmer tropical waters, likely owing to fish predators. Our results provide robust support for a temperature-dependent gradient in interaction strength and have potential implications for how marine ecosystems will respond to ocean warming.

The strength of species interactions, such as predation and competition, is thought to peak at low tropical latitudes and decline toward the poles (*1*). Such geographic variation in interaction strength is invoked frequently as both a major cause and consequence of the latitudinal diversity gradient, one of the most robust patterns of life on Earth (*2*–*5*). However, studies available to date across large spatial scales and multiple habitats provide conflicting support for increased predation intensity in the tropics and have been mostly limited to measuring rates of prey loss. For example, predation intensity (consumption rate) on seeds (*6*) and terrestrial insect mimics (*7*) was greater in the tropics than at higher latitudes. By contrast, attacks on open ocean long-line fishing hooks baited with natural prey peaked at mid-latitudes instead of the tropics (*8*), as did consumption of squid baits in shallow coastal waters (*9*).

Currently, it remains largely unknown whether global gradients in predation intensity produce associated gradients in the magnitude of effects on prey communities, especially across latitudes. Such a gradient in community-level effects is likely to have profound consequences for patterns of biodiversity (*10*), ecosystem function (*11*, *12*), and resilience to global change (*13*). Although some studies have found evidence for stronger effects of predation on community composition at tropical versus temperate sites, primarily in shallow-water marine benthic habitats (*14*–*17*), these were restricted to spatial scales of 20° to 45° latitude and usually along single coastlines. Other regional-scale studies in similar marine habitats did not detect this latitudinal pattern in community effects of predators (*18*, *19*). Where latitudinal trends in predation intensity and impact have been observed at regional spatial scales, a number of environmental factors that follow a latitudinal gradient have been proposed as drivers of this pattern, including time since glaciation, lack of freezing winters, day length, and temperature (*20*). Ambient temperature is likely important because it strongly influences metabolic rates and underpins organism functioning and the ecology of populations, communities, and ecosystems (*21*). Although temperature generally declines with latitude, the relationship varies among regions (Fig. 1). Thus, including in situ temperature as an independent predictor could help to explain the mixed results from previous studies. Clarifying the relationship between predation intensity, impacts on prey communities, and temperature could also facilitate prediction of community response to future ocean warming.

We tested whether intensity of predation and its community-level effects decrease from tropical to subpolar latitudes in coastal marine ecosystems. Specifically, we assessed the impact of fish and other large, mobile predators on sessile marine invertebrate communities. We used standardized and replicated experiments at *36* nearshore sites across 115° of latitude, along both Pacific and Atlantic coasts of the Americas (Fig. 1 and table S1). We conducted three complementary experiments to test whether predation intensity and top-down control of prey communities vary consistently along latitudinal and temperature gradients in both hemispheres. We focused on coastal subtidal communities of sessile invertebrates on hard substrates for multiple reasons. These communities are widely distributed throughout the world and are especially conducive to experiments, responding rapidly to manipulation and allowing for robust tests of general ecological processes (*3*, *22*). There is also evidence that top-down control is stronger in the tropics than in temperate regions for these hard-substrate communities at some regional scales (*14*–*16*, *18*, *23*). We expanded on this past work to test with high replication whether results are consistent on an extensive geographic scale, across the Americas in two oceans (*24*).

Our experiments measured three separate components of predation: (i) consumption of a standard bait as a measure of predation intensity, (ii) effects of sustained predation on the development of benthic community composition and biomass over 3 months, and (iii) the effects of short-term predation on already developed benthic communities (table S2) (*24*). The three complementary predation measures were colocated in space and time at each site. To compare predator consumption rates on a broadly palatable prey for the first component, we used dried squid as a standardized bait at all sites and recorded bait loss after 1 hour as a measure of predation intensity (*25*). For the second and third components, we allowed natural communities to develop on standardized substrates for 3 months (*15*) and manipulated predator access at different time points in community assembly, to evaluate the effect of predation on composition and biomass of sessile invertebrate communities (*24*). Cages were designed and used in both experiments to selectively exclude and evaluate effects of large (>1 cm) mobile predators, especially fishes, which are major consumers of benthic invertebrate prey in shallow subtidal habitats and can affect their community composition (*14*–*18*, *23*). The second component contrasted communities developed continuously under caged versus uncaged control conditions for

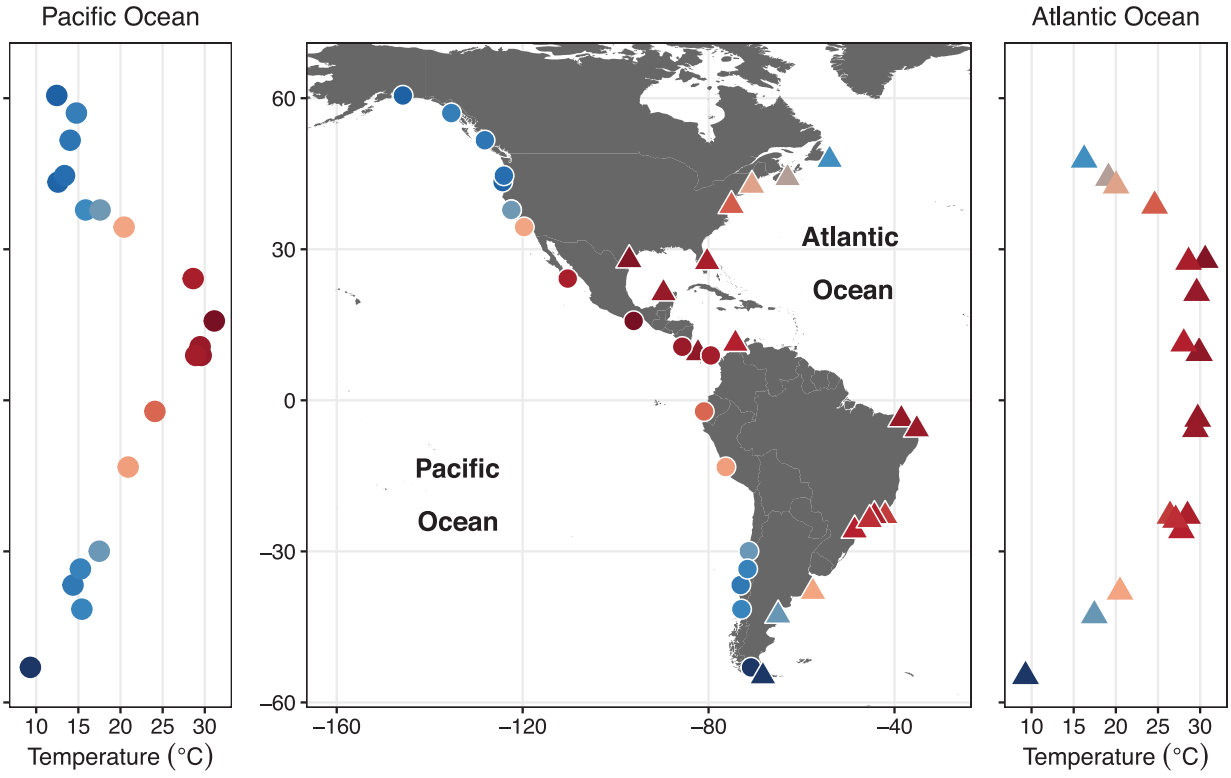


Fig. 1. Site location and mean temperatures. Location, latitude, and mean temperature recorded at experimental sites on Atlantic (triangle) and Pacific (circle) coastlines of the Americas. Color scale indicates gradient in temperature recorded across latitudes during the experiment (dark blue, ~9°C; dark red, ~31°C).

12 weeks. For the third component, we allowed communities to develop for 10 weeks in cages and then uncaged half of these, comparing effects of predator exposure on these established communities after 2 additional weeks. We also measured temperature at each site throughout the experiments using dataloggers (*24*).

We analyzed the results with mixed effects models and a model selection approach, with separate global models estimating the responses of bait consumption; sessile community biomass; and community composition to variation in seawater temperature or latitude, ocean basin, hemisphere, caging treatment, and interactions among all these terms. We explicitly compared alternate models that included either latitude or temperature recorded during the experiment to evaluate which was a better predictor of predator effects (*24*).

Our results provide robust experimental evidence that top-down control of community

[1]Smithsonian Environmental Research Center, Tiburon, CA and Edgewater, MD, USA. [2]Department of Biology, Temple University, Philadelphia, PA, USA. [3]Smithsonian Tropical Research Institute, Balboa, Ancon, Republic of Panama. [4]Tennenbaum Marine Observatories Network and MarineGEO program, Smithsonian Institution, Edgewater, MD, USA. [5]Marine Science Institute, University of California, Santa Barbara, CA, USA. [6]United States Naval Academy Oceanography Department, Annapolis, MD, USA. [7]Centro Austral de Investigaciones Científicas (CADIC-CONICET), Ushuaia, Tierra del Fuego, Argentina. [8]Department of Environmental Engineering Sciences, Engineering School of Sustainable Infrastructure and Environment, University of Florida, Gainesville, FL, USA. [9]Department of Marine Biotechnology, Instituto de Estudos do Mar Almirante Paulo Moreira, Arraial do Cabo, RJ, Brazil. [10]Laboratorio de Sistemática de Invertebrados Marinos (LABSIM), Universidad del Mar, campus Puerto Angel, Oaxaca, Mexico. [11]Instituto Patagónico para el Estudio de los Ecosistemas Continentales (IPEEC-CONICET), Puerto Madryn, Chubut, Argentina. [12]Departamento de Ecología, Facultad de Ciencias, Universidad Católica de la Santísima Concepción, Concepción, Chile. [13]Centro de Investigación en Biodiversidad y Ambientes Sustenables (CIBAS), Universidad Católica de la Santísima Concepción, Concepción, Chile. [14]Hakai Institute, Heriot Bay, BC, Canada. [15]School of Environmental Studies, University of Victoria, Victoria, BC, Canada. [16]Centro i-mar and CeBiB, Universidad de Los Lagos, Puerto Montt, Chile. [17]Mission-Aransas NERR, University of Texas Marine Science Institute, Port Aransas, TX, USA. [18]Charles Darwin Research Station, Charles Darwin Foundation, Santa Cruz, Galapagos, Ecuador. [19]Gloucester Marine Station, Department of Environmental Conservation, University of Massachusetts, Amherst, MA, USA. [20]Universidade Federal do ABC, São Bernardo do Campo, SP, Brazil. [21]Bedford Institute of Oceanography, Fisheries and Oceans Canada, Dartmouth, NS, Canada. [22]Centre for Marine Biology, University of São Paulo, São Sebastião, SP, Brazil. [23]Zoology Department, University Federal do Paraná, Curitiba, PR, Brazil. [24]Departamento de Ciencias Marinas y Costeras, Universidad Autónoma de Baja California Sur, La Paz, BCS, Mexico. [25]Departamento de Química y Biología, Universidad San Ignacio de Loyola, Lima, Peru. [26]Smithsonian Marine Station, Fort Pierce, FL, USA. [27]Laboratorio MARINAR, Facultad de Ciencias Biológicas, Universidad Nacional Mayor de San Marcos, Lima, Peru. [28]Asociacion Conservaccion, Lima, Peru. [29]Departamento de Ecología, Facultad de Ciencias, Universidad Católica de la Santísima Concepción, Concepción, Chile. [30]Centro FONDAP de Investigación de Dinámicas de Ecosistemas Marinos de Altas Latitudes (IDEAL), Chile. [31]Facultad de Ciencias Naturales e Ingeniería, Universidad Jorge Tadeo Lozano, Santa Marta, Colombia. [32]Departamento de Oceanografia e Limnologia, Federal University of Rio Grande do Norte, Brazil. [33]Departamento de Biologia, Universidade Federal do Ceará, Fortaleza, CE, Brazil. [34]Northwest Atlantic Fisheries Centre, Fisheries and Oceans Canada, St. John's, NL Canada. [35]Oregon State University, Coastal Oregon Marine Experiment Station, Newport, OR, USA. [36]Facultad de Ciencias del Mar, Universidad Católica del Norte, Coquimbo, Chile. [37]Millennium Nucleus Ecology and Sustainable Management of Oceanic Island (ESMOI), Coquimbo, Chile. [38]Centro de Estudios Avanzados en Zonas Áridas (CEAZA), Coquimbo, Chile. [39]Estación Costera de Investigaciones Marinas, Pontificia Universidad Católica de Chile, Las Cruces, Chile. [40]Posgrado en Ciencias del Mar y Limnología, Universidad Nacional Autónoma de México (UNAM), Ciudad de México, México. [41]Unidad Multidisciplinaria de Docencia e Investigación Sisal (UMDI-SISAL), Facultad de Ciencias, Universidad Nacional Autónoma de México (UNAM), Yucatán, México. [42]Argentino de Ciencias Naturales Bernardino Rivadavia, CONICET, Argentina. [43]Virginia Institute of Marine Science, College of William and Mary, VA, USA. [44]Oil Spill Recovery Institute/PWSSC, Cordova, AK, USA. [45]CIT Santa Cruz (CONICET-UNPA), IIMyC (CONICET-FCEyN, UNMdP), Argentina. [46]Instituto de Biología de Organismos Marinos (IBIOMAR-CONICET), Puerto Madryn, Chubut, Argentina. [47]Zukunft Umwelt Gesellschaft (ZUG) gGmbH, International Climate Initiative, Berlin, Germany. [48]University of Oregon, Oregon Institute of Marine Biology, Charleston, OR, USA. [49]Laboratorio Nacional de Resiliencia Costera (LANRESC), CONACYT, Sisal, Yucatan, Mexico. [50]International Chair for Ocean and Coastal Studies, Harte Research Institute, Texas A&M University at Corpus Christi (TAMUCC), Corpus Christi, Texas, USA. [51]Universidade do Estado do Rio de Janeiro, Brazil. [52]Instituto de Ciencias Marinas y Limnológicas, Facultad de Ciencias, Universidad Austral de Chile, Campus Isla Teja, Valdivia, Chile. [53]Center for Conservation and Sustainability, Smithsonian Conservation Biology Institute, National Zoological Park, Washington, DC, USA. [54]Hunt LNG Operating Company, Peru. [55]Centro de Investigación en Ciencias del Mar y Limnología (CIMAR), San José, Costa Rica. [56]Biodiversity Research Centre, University of British Columbia, BC, Canada. [57]University of Aberdeen, Oceanlab, Aberdeen, Scotland.
***Corresponding author. Email: ashtong@si.edu (G.V.A.); ruizg@si.edu (G.M.R.)**

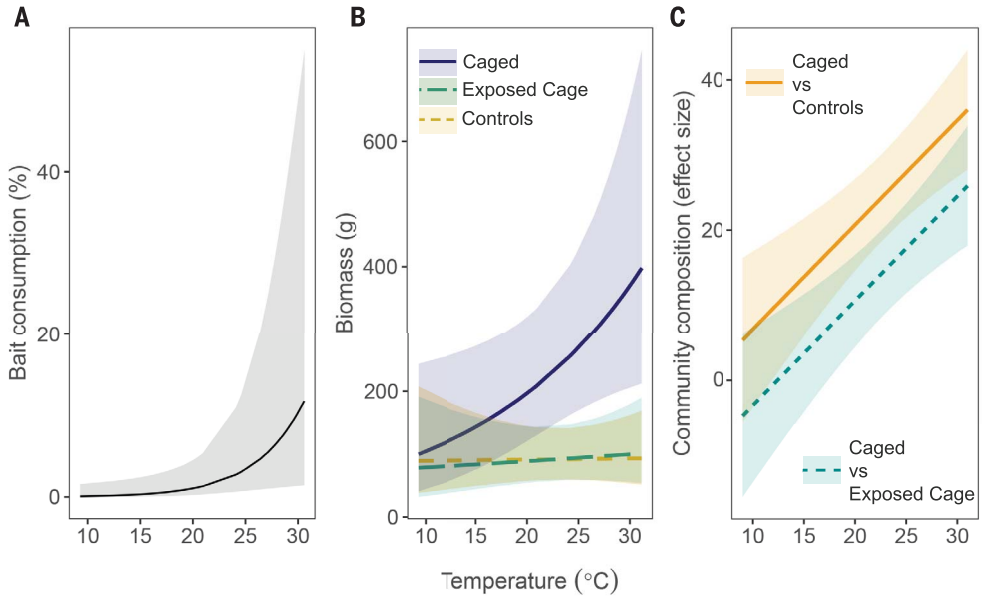


Fig. 2. Modeled variation in predation intensity and responses of biomass and community composition to predation with increasing temperature. (**A**) Predation measured as bait loss increased with in situ temperature along Atlantic and Pacific coastlines of the Americas. The line indicates predictions from a generalized linear mixed effects model [conditional coefficient of determination (R^2) = 0.79]. (**B**) The effect of predation on biomass accumulation increased with temperature. Dark blue indicates predators were excluded throughout the experiment; green indicates predators were excluded until the last 2 weeks of the experiment and then the experiment was exposed to predators; and yellow indicates open to predators throughout the experiment (model conditional R^2 = 0.89). Predators consumed significantly more biomass as temperature increased between 9° and 31°C. (**C**) Effect of predation on community composition increased along the latitudinal temperature gradient. Exclusion of predators throughout the 3-month experiment (gold, caged versus controls) had a greater impact on community composition than 2-week exposure (blue, caged versus exposed cage) of the late-stage community to predators. Lines show effect size as predictions from linear models of square roots of the estimated component of variation for each contrast within each site. Shaded areas show 95% confidence intervals (CIs) (*24*).

structure consistently increases with temperature and is strongest in the tropics, supporting a major tenet in ecology and evolutionary biology. Predation intensity and its effects on marine hard-substrate communities increased from colder high-latitude to warmer tropical waters (Fig. 2). Seawater temperature and latitude were strongly correlated [correlation coefficient (r) = 0.84], and although results were qualitatively similar for seawater temperature and absolute latitude, the models with seawater temperature were more strongly supported for both predation intensity and community responses (*24*). Predation intensity, as measured in the first experiment with bait consumption, was greatest in the warm tropics and approached zero at sites where mean summer sea surface temperature was below ~20°C (Fig. 2A, fig. S2, and table S3). Whereas the bait loss assay provides a short-term (1 hour) measure of predation intensity, the two caging experiments integrate longer-term impacts of predators on community attributes, revealing that predators had consistently larger effects on communities at higher temperatures and during multiple stages of community development. Specifically, in the second experiment, the effect of predators increased with temperature for both biomass accumulation (wet-weight) (Fig. 2B, fig. S3, and table S4) and community composition (Fig. 2C, figs. S4 to S6, and tables S5 to S7). In the third experiment, predators reduced prey community biomass in warmer tropical waters during the 2-week exposure, compared with communities that remained caged, and biomass of these exposed communities converged on uncaged control treatments across all temperatures (Fig. 2B and table S4). Community composition also responded more strongly to this later-stage predation at warmer sites (Fig. 2C and table S6). Thus, results of these three complementary experiments provide strong and consistent evidence that predation intensity by mobile predators is higher on average, and shapes community composition more strongly, in warm tropical waters.

The organisms that changed most in response to predators were solitary tunicates and encrusting bryozoans; dominance of these groups diverged among treatments with increasing temperature (fig. S4). At warm water sites, encrusting bryozoans were most prevalent on open control panels, whereas solitary tunicates occurred most frequently on caged panels that restricted predator access (Fig. 3 and table S7, C and D). This pattern may result from competitive release of less palatable bryozoans when spatially dominant tunicates are removed by predators during community assembly (*19*, *26*). When later-stage tropical communities were exposed to predators, solitary tunicate dominance was reduced (compared with caged panels), with a coincident increase in bare space (Fig. 3). Bare space decreased toward the tropics in all treatments. It is likely that prevalence of large solitary tunicates drove the observed higher biomass in treatments protected from predators at most sites (Fig. 2B).

Although we found a strong overall increase in predation intensity and top-down control at warmer temperatures, the scale of the responses varied among ocean basins and hemispheres. For example, bait loss and community composition responses were more marked in the northern hemisphere (figs. S2, A and B, and S6B), whereas the biomass response of prey communities was more apparent in the North Atlantic and South Pacific than other regions (fig. S3B). This variation likely derives from regional differences in the species and functional characteristics of predators and prey, environmental conditions other than temperature, and/or biological factors beyond those measured here (such as productivity) (*23*). Fundamental differences in oceanography exist at the ocean basin scale (for example, equatorial upwelling on the Pacific coastline is largely absent from the Atlantic sites) that would be expected to have effects on the observed latitudinal patterns (*27*). More broadly, the variation among sites underscores the need for high replication and broad geographic coverage to thoroughly evaluate both regional and global patterns.

This study provides new insights into the macroecological pattern of biotic interactions. We show that intensity of predation indeed declines consistently with latitude, as expected, but is better predicted by mean summer temperature experienced during the experiment than by latitude, hinting at underlying mechanisms. We demonstrate that this gradient in predation intensity produces a parallel gradient in top-down control of marine community

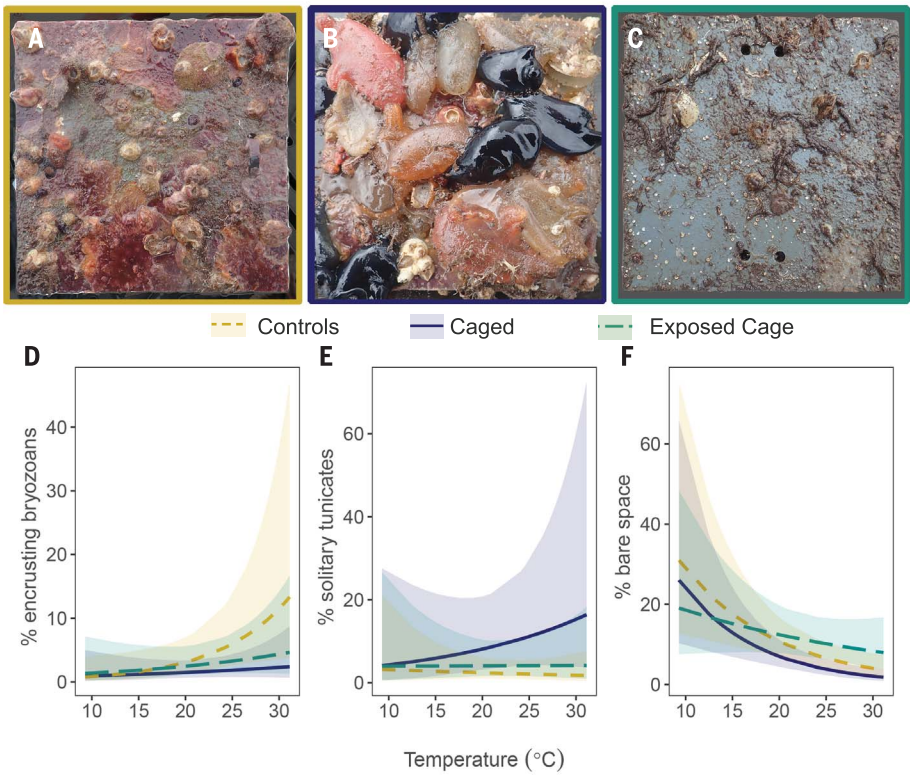


Fig. 3. Effects of predator treatments on community composition at a tropical Atlantic site and response of key functional groups from models based on all sites. (**A** to **C**) Photographs illustrate differences among experimental treatments at Bocas del Toro, Panama. At this and other warm water sites, encrusting bryozoans predominated in (A) control panels (exposed to predators), (B) solitary tunicates in caged panels (predators excluded), and (C) bare space in exposed cage panels [as in (B) but exposed to predators for the last 2 weeks through cage removal]. (**D** to **F**): Modeled percent cover across all sites of (D) encrusting bryozoans, (E) solitary tunicates, and (F) bare space, which together explained most of the variation in community composition among treatments (yellow, controls; dark blue, caged; green, exposed cage) in warm water sites. Shaded areas show 95% CIs (*24*).

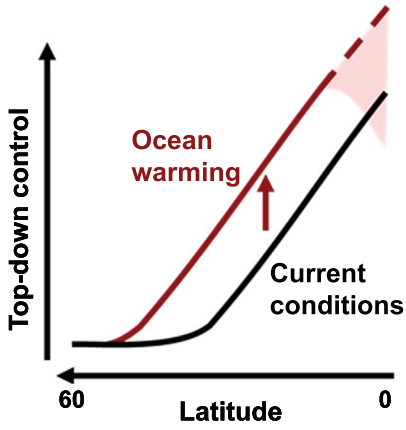


Fig. 4. Conceptual illustration of the hypothesized impact of ocean warming on future trends in top-down control of marine communities. Predation intensity was low and had little or no effect on benthic communities at cold latitudes and increased toward the equator with temperature, above an inflection point (~20°C). The black line describes a simplified view of the current latitudinal pattern of top-down control in our study. The solid red line describes the hypothesized effect of future ocean warming, which may shift this inflection point poleward, increasing predation effects at higher latitudes. The dashed red line describes a region of uncertainty in the tropics, where increased temperatures exceed our current observations and possibly thermal tolerance of some predators, so that top-down control may increase or decline within this region (shaded to suggest a range of possible responses).

biomass and composition that has been long suspected but not rigorously tested at this scale. As predicted, predation intensity in our shallow hard-substrate communities increased with temperature, similar to the patterns of bait loss in terrestrial and marine environments over an expansive latitudinal range (*7*, *9*). Our results were likely driven by highly mobile fish that can exert strong effects on epibenthic invertebrates in warm tropical water (*14*–*18*, *23*). We recognize that predation effects may differ for marine communities in other habitat types, including those where macroinvertebrates exert strong predation effects (*3*, *27*). More specifically, other studies in marine systems have shown a variety of patterns (*8*, *9*, *28*), which may reflect physical differences among habitats, taxonomic composition of predator or prey groups, smaller spatial scales, or less replication.

Overall, our analyses demonstrate a strong temperature-dependent gradient of increasing predator impacts on community biomass and composition and support prior predictions of stronger interaction strengths at warmer latitudes based on regional-scale studies [for example, (*15*, *17*)]. This study, completed at a large spatial scale, contributes to mounting evidence that temperature is a key predictor of global gradients, not only in diversity (*29*) and a suite of biological processes (*21*) but also in the strength of interactions among species (*30*, *31*) and the resulting effects of those interactions on communities.

Our results imply that climate change may have predictable effects on the regulation of nearshore communities along the world's shorelines. Our finding of a fundamental relationship between temperature and predation effects across large geographic scales suggests that, in addition to shifting species' distributions (*32*), ocean warming may cause the intensity of top-down control to expand poleward (Fig. 4). Specifically, the observed temperature-predation relationship exhibits an inflection point at ~20°C (Fig. 2) (*19*) that will likely move poleward with warming (Fig. 4), both promoting top-down control at high latitudes and increasing predation effects at mid- to high latitudes through time (*33*). The response to warming is less certain in the tropics, where predation may increase or decrease, because projected temperature increases are beyond our current range of observations and may exceed thermal tolerances of existing predators. Such broad-scale shifts in top-down control could have far-reaching consequences, given the key role of species interactions in maintaining ecosystem structure, diversity, biogeochemical processes, and the provision of critical ecosystem services to human communities (*3*, *13*).

REFERENCES AND NOTES

1. D. W. Schemske, G. G. Mittelbach, H. V. Cornell, J. M. Sobel, K. Roy, *Annu. Rev. Ecol. Evol. Syst.* **40**, 245–269 (2009).
2. H. Hillebrand, *Am. Nat.* **163**, 192–211 (2004).
3. R. T. Paine, *Am. Nat.* **100**, 65–75 (1966).
4. J. H. Connell, *Annu. Rev. Ecol. Syst.* **3**, 169–192 (1972).
5. G. J. Vermeij, E. Zipser, E. C. Dudley, *Paleobiology* **6**, 352–364 (1980).
6. A. L. Hargreaves *et al.*, *Sci. Adv.* **5**, eaau4403 (2019).
7. T. Roslin *et al.*, *Science* **356**, 742–744 (2017).
8. M. Roesti *et al.*, *Nat. Commun.* **11**, 1527 (2020).
9. M. A. Whalen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 28160–28166 (2020).
10. A. L. Freestone, B. D. Inouye, *Ecology* **96**, 264–273 (2015).
11. A. L. Freestone, G. M. Ruiz, M. E. Torchin, *Ecology* **94**, 1370–1377 (2013).
12. B. S. Cheng, A. H. Altieri, M. E. Torchin, G. M. Ruiz, *Ecology* **100**, e02617 (2019).

13. J. A. Estes *et al*., *Science* **333**, 301–306 (2011).
14. A. L. Freestone, R. W. Osman, G. M. Ruiz, M. E. Torchin, *Ecology* **92**, 983–993 (2011).
15. A. L. Freestone *et al*., *J. Anim. Ecol.* **89**, 323–333 (2020).
16. A. L. Freestone *et al*., *Ecology* **102**, e03428 (2021).
17. G. M. Dias *et al*., *Divers. Distrib.* **26**, 1198–1210 (2020).
18. L. P. Kremer, R. M. da Rocha, *Biol. Invasions* **18**, 3223–3237 (2016).
19. J. T. Lavender, K. A. Dafforn, M. J. Bishop, E. L. Johnston, *Ecology* **98**, 2391–2400 (2017).
20. A. Moles, J. Ollerton, *Biotropica* **48**, 141–145 (2016).
21. J. H. Brown, J. F. Gillooly, A. P. Allen, V. M. Savage, G. B. West, *Ecology* **85**, 1771–1789 (2004).
22. J. J. Stachowicz, R. B. Whitlatch, R. W. Osman, *Science* **286**, 1577–1579 (1999).
23. M. E. Torchin *et al*., *Ecology* **102**, e03434 (2021).
24. Materials and methods are available as supplementary materials.
25. J. E. Duffy, S. L. Ziegler, J. E. Campbell, P. M. Bippus, J. S. Lefcheck, *PLOS ONE* **10**, e0142994 (2015).
26. G. R. Russ, *Oecologia* **53**, 12–19 (1982).
27. K. Véliz, N. Chandía, K. Bischof, M. Thiel, *J. Phycol.* **56**, 1090–1102 (2020).
28. C. A. Musrri *et al*., *Mar. Biol.* **166**, 142 (2019).
29. D. P. Tittensor *et al*., *Nature* **466**, 1098–1101 (2010).
30. T. Wernberg *et al*., *Ecol. Lett.* **13**, 685–694 (2010).
31. R. L. Kordas, C. D. G. Harley, M. I. O'Connor, *J. Exp. Mar. Biol. Ecol.* **400**, 218–226 (2011).
32. E. S. Poloczanska *et al*., *Front. Mar. Sci.* **3**, 1–21 (2016).
33. V. Kubelka *et al*., *Science* **362**, 680–683 (2018).
34. G. V. Ashton *et al.*, Data for "Predator control of marine communities increases with temperature across 115 degrees of latitude". figshare (2022); https://doi.org/10.25573/serc.19469900.

ACKNOWLEDGMENTS

This paper is a product of the Pan American Experimental Initiative in Marine Macroecology (PanAmEx), supported by numerous people and grants since its inception. We are particularly indebted to field assistants, host marinas and ports, and institutions for their help in establishing and maintaining the sites. For their particular contributions we thank N. Abrain Sánchez, G. Agurto Rodríguez, P. Albuquerque, R. Altamirano, L. P. Alves, H. Galo Andrade, APPM Port Administration, M. Araya, A. Arnwine, C. Arriola, K. Bachen, M. Badillo-Aleman, J. Barley, H. Bartsch, N. Battini, L. Bent-Hooker, A. Beylan, J. Bleuel, E. Bobadilla, X. Boza, D. Branson, J. Bucholz, J. Bueno, A. Bungay, J. P. Carvalho Lima, Y. N. Casanova Salazar, M. C. Castellanos, R. Castillo, K. Castro, C. Cesar, I. Chacón, Club Náutico AFASyN, Club Nautico Mar del Plata, F. Contrera, M. Correal, L. Simioni Costa, C. Cruz-Gomez, K. Curiston, I. Davidson, Y. Davila, M. De Koster (Maritime Gloucester), D. de Miranda Lins, L. de Moura Oliveira, M. Lucila Del Gener, C. Détrée, A. Dias Kassuga, S. Diaz, R. DiMaria, R. Dueñas, K. Escalante-Herrera, ESPOL Polytechnic University, L. Falsone, J. Fedex, D. Fernandez Barboza, M. Ferreira Valença, B. Figueroa Soler, O. Florentin, L. Freyre, A. Fudge, Galapagos Biosecurity Agency, C. Giachetti, A. Giamportone, J. Gómez-Vásquez, J. Gonzalez, M. G. Vazquez, P. Guadarrama, E. Guerra, C. Guerra, J. Hardee, M. Harrigan Garfias (API, Huatulco), S. Havard, Y. A. Hernández, M. Hessing-Lewis, N. Hitchcock, T. Huber, I. Clube de Natal, K. Inagaki, Instituto Nacional de Pesca, Inversiones Marina Turística S.A., V. Jenkins, M. Kronman (Santa Barbara Harbor), M. Lacerda (Porto Real Marina), H. Lambert, K. Larson, A. Leduc, G. Lima, B. Lonzetti, F. Lopes Penha, E. Macaya, S. Martínez, K. Matheson, J. Medeiros de Oliveira, T. Mendoza, R. Menezes, H. F. Messano, M. Mews, M. Minton, K. Mitchell, J. Monteiro, A. Morales (API, México), T. Mullady, T. Murphy, K. Newcomer, S. Obenat, E. Oliveira, N. Ortiz, M. Oxxean, Ê. V. Paiva Bandeira, T. Palyo, L. Pardo, S. Pegau, N. Bonnemasou Peixoto, T. Pereira Menezes, R. Periera Menezes, J. Pinochet, C. Prentice, A. Ramos-Morales, M. Ramos-Sánchez, C. Reveles, V. Reyes, A. Reynolds, M. Saldaña, Salinas Yacht Club, J. P. Sánchez-Ovando, S. A. Santa-Cruz, K. Savage, F. Silva, I. Silva-Morales, C. Simkanin, G. Smith, S. Soria, C. Soto-Balderas, A. Sousa Matos, D. Sparks, A. Tissot, M. Torres, K. Treiberg (Santa Barbara Harbor), D. Ugalde, M. Ureña, L. Vallejos, D. Van Maanen, R. van Velzen, J. Vega-Sequeda, M. Vegh, R. B. Vera, J. J. Vera Duarte, M. Vergotti, M. Vieira da Silva, A. Villeneuve, W. Wied, T. Wells, R. Whippo, N. Williams, and Yacht Club de Ilhabela. We also thank the reviewers for their contribution to improving this manuscript for publication. **Funding:** This research has been supported by multiple grants and institutions, including Smithsonian Institution Hunterdon and Johnson Fund to G.M.R. and J.E.D.; NSF OCE grant 1434528 to A.L.F., G.M.R., and M.E.T.; financial support from FAPESP 2016/17647-5 and CNPq 308268/2019-9 to G.M.D.; Galapagos Conservancy, Lindblad Expedition/National Geographic Fund, Galapagos Conservation Trust, Paul M. Angell Foundation and Ecoventura, and the Charles Darwin Foundation to I.K.; Fisheries and Oceans Canada, Aquatic Invasive Species Science Program to C.H.M.; ANID-FONDECYT (# 1180647) and CeBiB (FB-0001) to A.H.B.; CONICYT-FONDECYT #1190529 and FONDAP #15150003 (IDEAL) to N.V.; ANID (ICN2019_015 and NCN19_05) to S.A.N.; Chilean Millennium Scientific Initiative (ESMOI), Chile, to M.M. and M.T.; ANID-FONDECYT #1190954 to M.T.; the Tula Foundation to M.W. and N.B.; CONICET-PIP 20130100508, ANPCyT-PICT 2016-1083 to E.S. and A.B.; CONICET, PIP 2018-2020. 11220170100643CO to M.A.; CAPES (Finance Code 001) to E.A.V.; FAPESP 2018/11044-2 to A.A.V.F.; CNPq 301601/2016-6 to A.A.V.F.; CNPq 309295/2018-1 to R.M.R.; Serrapilheira Institute (Serra-1708-15364) and CNPq (310517/2019-2) to G.O.L.; CONACyT 2018-000012-01NACF-08376 to L.A.P.-A.; CONACyT 2019-000002-01NACF-14266 to N.Y.S.-M.; and the Smithsonian Institution through the Tennenbaum Marine Observatory Network. This publication is contribution number 2335 of the Charles Darwin Foundation for the Galapagos Islands and contribution number 103 of the Tennenbaum Marine Observatories Network and MarineGEO Program. **Author contributions:** G.V.A., A.L.F., J.E.D., G.M.R., and M.E.T. conceived the project and design; G.V.A. and B.T. coordinated data collection; G.V.A. and B.J.S. analyzed the data; G.V.A., J.E.D., A.L.F., G.M.R., B.J.S., and M.E.T. wrote the manuscript drafts; all authors collected and/or helped collect field data. All authors commented on and/or approved the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data from this work can be found at (*34*). **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.abc4916
Materials and Methods
Figs. S1 to S6
Tables S1 to S7
References (*35*–*50*)
MDAR Reproducibility Checklist

Submitted 25 November 2020; resubmitted 16 September 2021
Accepted 3 May 2022
10.1126/science.abc4916

MICROBIOME

Robust variation in infant gut microbiome assembly across a spectrum of lifestyles

Matthew R. Olm[1]†, Dylan Dahan[1]†, Matthew M. Carter[1], Bryan D. Merrill[1], Feiqiao B. Yu[2], Sunit Jain[2], Xiandong Meng[3], Surya Tripathi[4], Hannah Wastyk[1], Norma Neff[2], Susan Holmes[1,5], Erica D. Sonnenburg[1], Aashish R. Jha[6], Justin L. Sonnenburg[1,2]*

Infant microbiome assembly has been intensely studied in infants from industrialized nations, but little is known about this process in nonindustrialized populations. We deeply sequenced infant stool samples from the Hadza hunter-gatherers of Tanzania and analyzed them in a global meta-analysis. Infant microbiomes develop along lifestyle-associated trajectories, with more than 20% of genomes detected in the Hadza infant gut representing novel species. Industrialized infants—even those who are breastfed—have microbiomes characterized by a paucity of *Bifidobacterium infantis* and gene cassettes involved in human milk utilization. Strains within lifestyle-associated taxonomic groups are shared between mother-infant dyads, consistent with early life inheritance of lifestyle-shaped microbiomes. The population-specific differences in infant microbiome composition and function underscore the importance of studying microbiomes from people outside of wealthy, industrialized nations.

The human gut microbiome undergoes a complex process of assembly beginning immediately after birth (*1*). New microbes attempting to engraft within this community often depend upon niches established by previous colonizing species and thus the final adult microbiome composition may be contingent upon the species acquired early in life. The microbiome assembly process of infants living in industrialized nations is well studied and tends to follow a series of characterized steps that lead to the low-diversity gut microbiome composition characteristic of industrialized adults (*2*). The microbiome assembly process that occurs in infants living nonindustrialized lifestyles (which results in the characteristically diverse adult microbiomes of nonindustrialized adults) (*3*) is largely unknown (*4*). Of particular interest are the following: the timing at which the microbiomes of infants from different lifestyles diverge, the microbes and functions that are characteristic of infants from different lifestyles, and whether there are differences in the taxa that are vertically transmitted from mothers to infants, which seed the microbiome assembly process.

To address these questions we performed metagenomic sequencing on infant fecal samples from the Hadza, a group of modern hunter-gatherers in sub-Saharan Africa (*5*, *6*). The Hadza inhabit seminomadic bush camps of ~5 to 30 people, exhibit a moderate level of community child rearing within these camps (*7*), and are breastfed early in life and weaned onto a diet of baobab powder and premasticated meat at ~2 to 3 years of age (*8*, *9*). In this study we (i) curated and analyzed a global dataset of 1900 16S rRNA sequencing samples of healthy infant fecal samples from 18 populations (including 62 Hadza infant samples) (*2*, *3*, *5*, *10*–*14*) to contextualize the Hadza infant microbiome (figs. S1 and S2), and (ii) performed deep metagenomic sequencing on 39 Hadza infant fecal samples and corresponding maternal fecal samples for 23 infants in order to assess subspecies variation, functional potential, and patterns of vertical transmission (tables S1 and S2).

A UniFrac ordination created from all 1900 16S rRNA sequencing samples revealed age and lifestyle to be strongly associated with the first and second axes of variation, respectively (Fig. 1A) (EnvFit; n = 1900; R^2 = 0.43 and 0.50; P = 0.001 and 0.001). Comparing populations that practice different lifestyles within the same country demonstrates that shared lifestyle affects microbiota composition more than geographic proximity (Fig. 1A, right panel, and fig. S3). The microbiome of infants living industrialized lifestyles diverges from others within the first 6 months of life, whereas the microbiomes of infants living transitional versus nonindustrialized lifestyles diverge at ~30 months of life (Fig. 1B). DNA extraction methods, differences in feeding practices, or other study-specific aspects may contribute to some of the variation in data. Intermediate trajectories are exhibited by populations on the boundaries of industrialized or nonindustrialized lifestyles (Fig. 1B, dashed lines), highlighting the apparent sensitivity of infant microbiota development to lifestyle-related factors.

We identified five microbial coabundance groups (CAGs) (*15*, *16*) in our dataset, which together account for an average of 93.8% of the microbiota composition per sample (Fig. 1C and fig. S4). The *Bifidobacterium-Streptococcus* CAG dominates infants from all lifestyles in early life (0 to 6 months), and over time this CAG yields to the *Bacteroides-Ruminococcus gnavus* CAG in industrialized infants and the *Prevotella-Faecalibacterium* CAG in infants living transitional or nonindustrialized lifestyles (Fig. 1C). Lifestyle-related differences in dominant CAGs become more pronounced over time and mirror taxonomic trade-offs observed in late infancy (*17*) and adulthood (*5*).

We next used our deep metagenomic sequencing data to assess microbiome-encoded functional differences between lifestyles. Broad lifestyle and age associated differences are seen in the overall functional capacity of the infant microbiomes (Fig. 2A), consistent with 16S rRNA amplicon-based analysis (Fig. 1A). Hadza infant metagenomes were assembled and binned into metagenome-assembled genomes (MAGs) representing 745 species, 175 (23.4%) of which represent novel species compared to the Unified Human Gastrointestinal Genome (UHGG) collection (*18*) (table S3). Novel species were recovered from diverse phylogenetic groups (fig. S5A); 88.6% (n = 155) were recovered from multiple Hadza samples (fig. S5B) and their genome quality was observed to be similar to that of genomes in the UHGG (fig. S5C). To assess prevalence through read mapping, MAGs were integrated with genomes recovered from Hadza adults (*19*) and public genomes from the human gut (*18*) into a comprehensive database of 5755 species-representative genomes. Overall, 23.4% of microbial species detected in the Hadza infants represent novel species (table S4). These data support that—similar to the adult Hadza gut—the Hadza infant gut contains extensive previously uncharacterized diversity.

The taxonomic specificity afforded by metagenomic sequencing allowed us to identify particular species that are depleted or enriched in infants living industrialized versus nonindustrialized lifestyles. Identified among the infants in this analysis were 310 VANISH (Volatile and/or Negatively associated in Industrialized Societies of Humans) and 12 BloSSUM (Bloom or Selected in Societies of Urbanization/Modernization) species (table S5 and fig. S6). Comparison against a large database of microbial species from nonhuman habitats (*20*) revealed that no VANISH and only one BloSSUM species match genomes recovered outside of the digestive tract or industrial wastewater, whereas 21 VANISH and three BloSSUM species match microbes recovered from nonhuman animal feces (table S6). VANISH species are more numerous and abundant than BloSSUM (fig. S7), and 63 VANISH species are effectively extinct (never detected) in infants living industrialized or transitional lifestyles. Many VANISH species (45.2%; 140 of 310) are

[1]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA. [2]Chan Zuckerberg Biohub, San Francisco, CA, USA. [3]ChEM-H Institute, Stanford University, Stanford, CA 94305, USA. [4]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, USA. [5]Department of Statistics, Stanford University, Stanford, CA, USA. [6]Genetic Heritage Group, Program in Biology, New York University Abu Dhabi, Abu Dhabi,United Arab Emirates.
*Corresponding author. jsonnenburg@stanford.edu
†These authors contributed equally to this work.

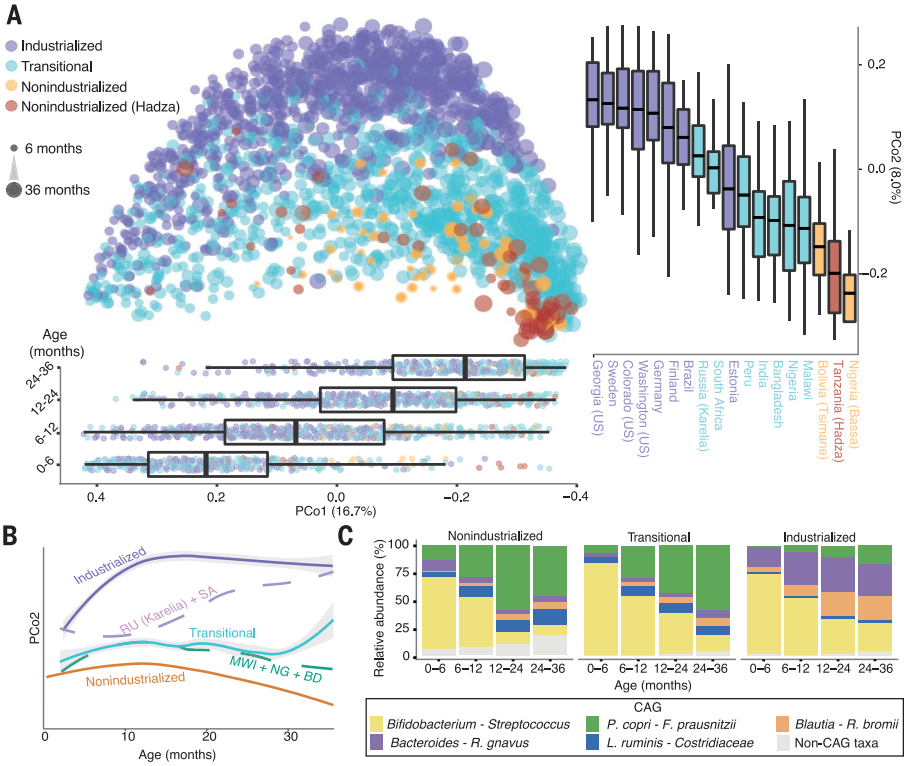


Fig. 1. Age and lifestyle are associated with infant microbiome composition. (**A**) Unweighted UniFrac dissimilarity Principal Coordinates Analysis (PCoA) (top left panel) of 1900 fecal samples from infants (<3 years old) across 18 populations based on amplicon sequence variant abundance. Point color indicates lifestyle and point size is proportional to age in months. Boxplots show the distribution of indicated age groups along PCo1 (bottom) and cohorts along PCo2 (right). (**B**) PCo2 versus sample age for the three lifestyle categories (solid lines) and specific indicated subpopulations (dashed lines). The purple dashed line includes Russia (Karelia) and South Africa [RU (Karelia) + SA] and the green dashed line includes Malawi, Nigeria (Urban), and Bangladesh (MWI + NG + BD). The middle transitional line (blue) contains all transitional samples. Lines are the smoothed conditional mean of PCo2 loadings (loess fit). (**C**) Relative abundance of CAGs by age group and lifestyle. Taxa in annotation are the most abundant taxa in a CAG.

present at 0 to 6 months in nonindustrialized infants whereas BloSSUM species are rarely detected this early in industrialized lifestyle infants (16.7%; 2 of 12) (Fig. 2B). Together these patterns suggest that more species are lost than gained as lifestyles industrialize.

Amplicon and metagenomic data both show that *Bifidobacterium* is the most prevalent taxon in early life (Figs. 1C and 2B). In the first 6 months, infants living nonindustrial lifestyles are dominated by *Bifidobacterium infantis* (also known as *Bifidobacterium longum* subsp. *infantis*) (Fig. 2C), a prolific utilizer of human milk oligosaccharides (HMOs) that is positively associated with human health and commonly used in probiotic supplements (*21*). *B. infantis* is significantly depleted in industrial microbiomes at 0 to 6 months ($P = 0.04$; $n = 151$ industrialized infants; $n = 27$ nonindustrial infants; Wilcoxon rank-sum test) and found at intermediate levels in transitional infants (Fig. 2C). *Bifidobacterium breve*, a species capable of limited HMO degradation (*22*), is instead the most abundant *Bifidobacterium* species in industrialized infants (Fig. 2C). *B. infantis* is antiassociated with *B. breve* in infants across all lifestyles (Fig. 2D). This trend also holds specifically among industrialized infants (correlation = −0.41, $P = 1.0 \times 10^{-3}$, $n =$ 62 industrialized infants, Spearman two sided hypothesis test), suggesting it may be driven by competitive exclusion rather than lifestyle-specific factors.

To determine whether these species-level differences result in community-wide differences in HMO degradation capacity, we mapped our metagenomic reads to the most well-characterized genetic clusters for human milk utilization (table S7). Five of these clusters are involved in HMO degradation (H1 to H5) and one is involved in nitrogen scavenging (referred to as the "urease" cluster) (*21*, *23*); recent studies have linked their expression in the infant gut microbiome to systemic immunological health outcomes (*24*). Five of the six clusters are more prevalent in nonindustrialized than industrialized infants, and their prevalence among transitional infants occurs between these two extremes (Fig. 2E). The H5 cluster, however, exhibits continued persistence beyond the first year of life only in infants from industrialized lifestyles (Fig. 2E). The H5 cluster encodes an ABC-type transporter known to bind core HMO structures, and it is more commonly found in *B. breve* than *B. infantis* (present in 119 of 129 *B. breve* MAGs and 41 of 69 *B. infantis* MAGs recovered from industrialized infants; $P = 1.4 \times 10^{-9}$, Fisher's exact test). The persistence of the H5 cluster beyond 12 months in industrialized infants—a time period in which breastfeeding is less common in these populations—suggests this cassette of genes exists in genomes that are not reliant upon breastfeeding. Breast milk consumption among industrialized infants reduces—but does not eliminate—lifestyle-associated differences in *B. infantis* and HMO-degradation cassette prevalence (fig. S8).

We next investigated strain-level differences among *B. infantis* genomes recovered from infants aged 0 to 1 years old ($n = 96$ MAGs). Several lifestyle-associated functional differences were discovered including (i) enrichment of glycoside hydrolase family 163 (GH163), a CAZyme involved in the utilization of complex N-glycans (including those found on immunoglobulins), in nonindustrialized versus industrialized infants (*25*) (fig. S9, A and B), (ii) differential prevalence of three Pfams (including one related to flagellar assembly) (fig. S9C), and (iii) increased prevalence of four uncharacterized gene clusters in MAGs from nonindustrialized versus industrialized infants (fig. S9D). To verify these metagenomics-based findings, we isolated and sequenced 20 *B. infantis* strains from the same Hadza infant fecal samples (table S3). GH163 and all four gene clusters also showed enrichment among Hadza *B. infantis* isolates as compared to the public reference genomes (fig. S9). Finally, strong lifestyle-specific phylogenetic clustering was observed among *B. infantis* isolate sequences and MAGs (Fig. 2F). This observation of strong region-specific phylogenetic signals could reflect long-term, multigenerational vertical transmission (*26*).

To assess the extent of vertical strain transmission in the Hadza infants, we deeply sequenced fecal samples from corresponding Hadza mothers ($n = 23$ Hadza dyads). Detailed strain-tracking analysis was performed with inStrain (*27*) with a threshold for identical strains of 99.999% popANI (table S8). Dyad pairs share far more strains (6.4 versus 0.3)

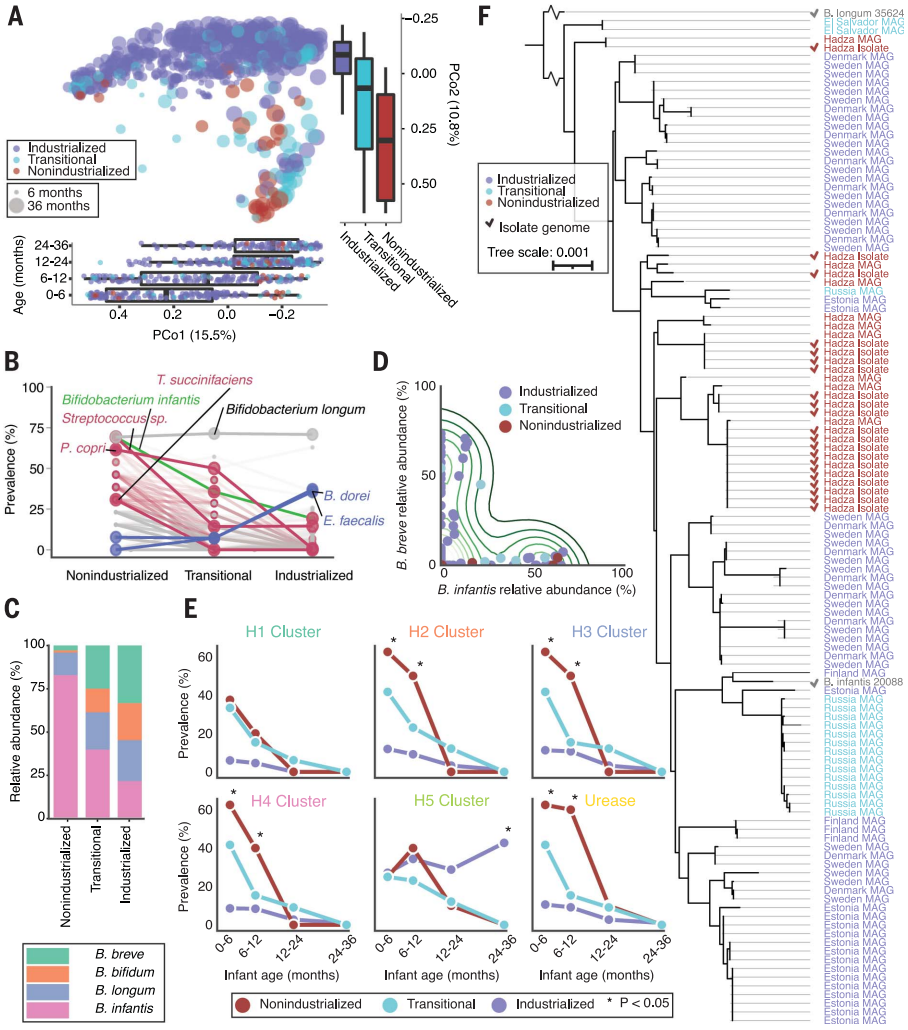


Fig. 2. Age and lifestyle are associated with infant microbiome functions. (**A**) PCoA on the basis of on 682 infant fecal metagenomes described at the gene abundance level in reads per kilobase million (RPKM). Points are colored by lifestyle and point size indicates infant age in months. Boxplots (bottom) show the distribution of indicated age groups in months along PCo1. Boxplots (right) show the distribution of each lifestyle along PCo2. The main axis of variation in this gene-based ordination is significantly associated with age (EnvFit; R^2 = 0.30; n = 679; P = 0.001) and the second axis of variation is significantly associated with lifestyle (EnvFit; R^2 = 0.35; n = 679; P = 0.001). (**B**) Prevalence of species across lifestyles among infants 0 to 6 months old. VANISH (red and green) and BloSSUM (blue) species with the lowest adj-P values have text annotations. *B. infantis* is shown in orange. "Other" taxa (gray) are those that do not significantly differ according to lifestyle. (**C**) Relative representation of four common *Bifidobacterium* species in infants 0 to 6 months old by lifestyle. (**D**) Scatterplot of *B. infantis* versus *B. breve* abundance among infants 0 to 6 months old. Contour lines display the kernel density estimation. (**E**) Prevalence of HMO-utilization clusters across ages and lifestyles. Clusters are considered present if all genes in the cluster are detected above a variable coverage threshold (to ensure that results are robust to differences in sequencing depth; see methods for details). * indicates adj-P < 0.05; Fisher's exact test with false discovery rate correction; nonindustrialized versus industrialized. (**F**) Phylogenetic tree of *B. infantis* genomes based on universal single copy genes. Genome names are colored on the basis of lifestyle of origin. Isolate genomes are marked with a checkmark. Public reference genomes for *B. longum* and *B. infantis* are included (gray text).

and have a higher percentage of strains shared (12.4% versus 0.5%) than nondyad pairs on average (P < 0.01, Wilcoxon rank-sum test) (Fig. 3A). Further, Hadza nondyads living in the same bush camp share more strains than those living in different bush camps (Fig. 3A) (P < 0.01, Wilcoxon rank-sum test), consistent with previously reported increased rates of strain sharing within Fijian social networks (*28*). Vertical strain sharing was detected among a range of phyla in the Hadza (Fig. 3B) and was higher among Bacteroidota and Cyanobacteria and lower among Firmicutes (Fisher's exact test with false discovery rate correction). Industrialized infants also exhibited increased and decreased vertical strain sharing of Bacteroidetes and Firmicutes, respectively (*29*). These results suggest that community interaction during rearing of infants and/or bush camp microenvironments may propagate group microbial sharing (*30*).

The same detailed strain-tracking analysis was next performed on a comparative dataset of 100 dyads from Sweden (*31*). Swedish and Hadza infants were 1.01 ± 0.00 and 0.95 ± 0.21 years old, respectively (P = 0.04, Wilcoxon rank-sum test); in addition, Swedish mothers were sampled immediately after birth whereas Hadza mothers were sampled contemporaneously with infants. Swedish infants born via C-section were excluded from this analysis (n = 17 eliminated) and in silico rarefaction was performed to account for differences in sequencing depth between the studies. Just as *Prevotella* and *Bacteroides* are enriched in nonindustrialized and industrialized infants, respectively (Fig. 1C), *Prevotella* and *Bacteroides* strains are more commonly vertically shared in Hadza and Swedish dyads, respectively (Fig. 3C; Fisher's exact test; P < 0.01). Similar trends are observed for VANISH and BloSSUM taxa (Fig. 3C). The species more abundant in maternal samples were more likely to be vertically transmitted (fig. S10); however, the small difference in infant age between populations may contribute to some differences. The findings suggest that vertical transmission may be a mechanism by which microbiota change is propagated over generations in response to altered lifestyles (*32*–*34*).

Taken together, our data show that infants from all lifestyles begin life with similar *Bifidobacteria*-dominated gut microbiota compositions, but subtle differences detected in early life compound over time. Differences in the species composition and HMO-degradation genes of the initially dominant *Bifidobacterium* communities are especially relevant as recent studies of these same genes suggest that their depletion in industrialized infants could have long-term negative immune consequences (*24*). The same taxa that differentiate lifestyles at 0 to 6 months of life are those that are most commonly vertically transmitted, suggesting that vertical transmission may help establish alternative development trajectories. Crucially, infants living transitional lifestyles display intermediate phenotypes between those of industrialized and nonindustrialized infants in almost all analyses performed. Although not conclusive, this is an important piece of evidence pointing to lifestyle as a possible causative factor in infant microbiome assembly. The Hadza-specific discoveries reported in this work (including the finding of increased nondyad

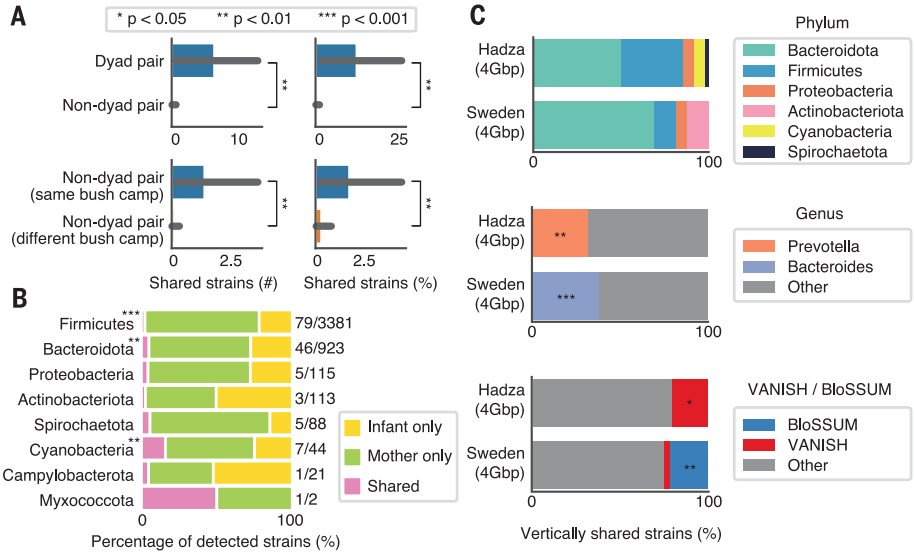


Fig. 3. Strain sharing between mother-infant dyads and nondyads is lifestyle-specific. (**A**) The mean strains shared (left) and the percentage of infant strains found in mothers (right) in mother-infant dyads versus mother-infant nondyads (top) and nondyads from the same bushcamp versus nondyads from different bushcamps (bottom). Error bars represent standard error (*, adj-P < 0.05; **, adj-P < 0.01; ***, adj-P < 0.001; Wilcoxon rank-sum test). (**B**) Percentage of strains detected in all Hadza mothers and infants and whether they are detected in infants only, mothers only, or shared within a mother-infant dyad ("shared") categorized by phylum. Numbers to the right of bars indicate the number of vertically shared strains over the number of strains detected in either infant or maternal samples. Phyla with a significant difference in the percentage of vertically transmitted strains as compared with all other phyla are marked with asterisks (Fisher's exact test with P value correction). (**C**) Percentage of vertically transmitted strains in Hadza and Swedish cohorts by phylum (top), genus (middle; only genera with significant differences shown), and VANISH / BloSSUM (bottom). All metagenomes were subset to 4Gbp for this analysis to remove any biases associated with sequencing depth. Taxa that are significantly enriched in either cohort are marked with an asterisk (*, adj-P < 0.05; **, adj-P < 0.01; ***, adj-P < 0.001; Fisher's exact test).

vertical transmission among members of the same bush camp, a social structure with no equivalent among industrialized communities) exemplify the importance of studying people outside of industrialized nations and highlights the need for additional studies to provide equity in understanding microbiomes across global societies. Our results also highlight the question of whether lifestyle-specific differences in the gut microbiome's developmental trajectory predispose populations to diseases common in the industrialized world, such as those driven by chronic inflammation (*35*, *36*).

REFERENCES AND NOTES

1. T. Yatsunenko *et al.*, *Nature* **486**, 222–227 (2012).
2. C. J. Stewart *et al.*, *Nature* **562**, 583–588 (2018).
3. G. K. Fragiadakis *et al.*, *Gut Microbes* **10**, 216–227 (2019).
4. M. C. de Goffau *et al.*, *Nat. Microbiol.* **7**, 132–144 (2022).
5. S. A. Smits *et al.*, *Science* **357**, 802–806 (2017).
6. F. Marlowe, *The Hadza: Hunter-Gatherers of Tanzania* (Univ. of California Press, 2010).
7. N. G. Blurton Jones *et al.*, *Am. J. Phys. Anthropol.* **89**, 159–181 (1992).
8. A. N. Crittenden, N. L. Conklin-Brittain, D. A. Zes, M. J. Schoeninger, F. W. Marlowe, *Evol. Hum. Behav.* **34**, 299–304 (2013).
9. A. N. Crittenden, F. W. Marlowe, *Hum. Nat.* **19**, 249–262 (2008).
10. A. S. Raman *et al.*, *Science* **365**, eaau4735 (2019).
11. T. Vatanen *et al.*, *Cell* **165**, 842–853 (2016).
12. F. A. Ayeni *et al.*, *Cell Rep.* **23**, 3056–3067 (2018).
13. A. W. Kamng'ona *et al.*, *Sci. Rep.* **9**, 12893 (2019).
14. D. D. Sprockett *et al.*, *Nat. Commun.* **11**, 3772 (2020).
15. S. C. Watts, S. C. Ritchie, M. Inouye, K. E. Holt, *Bioinformatics* **35**, 1064–1066 (2019).
16. A. R. Jha *et al.*, *PLOS Biol.* **16**, e2005396 (2018).
17. J. Roswall *et al.*, *Cell Host Microbe* **29**, 765–776.e3 (2021).
18. A. Almeida *et al.*, *Nat. Biotechnol.* **39**, 105–114 (2021).
19. B. D. Merrill *et al.*, *bioRxiv* (2022), p. 2022.03.30.486478.
20. S. Nayfach *et al.*, *Nat. Biotechnol.* **39**, 499–509 (2021).
21. R. M. Duar *et al.*, *Nutrients* **12**, 3247 (2020).
22. M. Sakanaka *et al.*, *Nutrients* **12**, 71 (2019).
23. R. G. LoCascio, P. Desai, D. A. Sela, B. Weimer, D. A. Mills, *Appl. Environ. Microbiol.* **76**, 7373–7381 (2010).
24. B. M. Henrick *et al.*, *Cell* **184**, 3884–3898.e11 (2021).
25. J. Briliūtė *et al.*, *Nat. Microbiol.* **4**, 1571–1581 (2019).
26. P. Ferretti *et al.*, *Cell Host Microbe* **24**, 133–145.e5 (2018).
27. M. R. Olm *et al.*, *Nat. Biotechnol.* **39**, 727–736 (2021).
28. I. L. Brito *et al.*, *Nat. Microbiol.* **4**, 964–971 (2019).
29. Y. C. Lou *et al.*, *Cell Rep. Med.* **2**, 100393 (2021).
30. A. H. Moeller *et al.*, *Science* **353**, 380–382 (2016).
31. F. Bäckhed *et al.*, *Cell Host Microbe* **17**, 690–703 (2015).
32. E. D. Sonnenburg *et al.*, *Nature* **529**, 212–215 (2016).
33. P. Vangay *et al.*, *Cell* **175**, 962–972.e10 (2018).
34. M. J. Blaser, *Cell* **172**, 1173–1177 (2018).
35. J. L. Sonnenburg, E. D. Sonnenburg, *Science* **366**, eaaw9255 (2019).
36. E. D. Sonnenburg, J. L. Sonnenburg, *Nat. Rev. Microbiol.* **17**, 383–390 (2019).

ACKNOWLEDGMENTS

We acknowledge the numerous people and organizations who provided logistical support and conducted sample collection in the USA, Tanzania, and Nepal, including Dorobo Safaris, the Human Food Project, J. Changalucha, A. Manjurano, M.G. Domiguez-Bello, M. St. Onge, A. Weakly, and Y. Gautam. We thank D. Relman and C. Damman for helpful discussion and input throughout project conceptualization and analysis. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This research utilizes data obtained by the TEDDY study group, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the National Institute of Allergy and Infectious Diseases (NIAID), the National Institute of Child Health and Human Development (NICHD), the National Institute of Environmental Health Sciences (NIEHS), Centers for Disease Control and Prevention (CDC), and JDRF. The data from the TEDDY study reported here were supplied by the database of Genotypes and Phenotypes (dbGaP; Study Accession: phs001443.v1.p1), which is maintained by the National Center for Biotechnology Information (NCBI). This manuscript was not prepared in collaboration with investigators of the TEDDY study and does not necessarily reflect the opinions or views of the TEDDY study, dbGaP, or the NIDDK. J.L.S. is a Chan-Zuckerberg Biohub Investigator. **Recognition of work on Indigenous communities:** Research involving Indigenous communities is needed for a variety of reasons, including assurance that scientific discoveries and understanding appropriately represent all populations and do not only benefit those living in industrialized nations. Special consideration must be made to ensure that this research is conducted ethically and in a nonexploitative manner. In this study we performed deep metagenomic sequencing on fecal samples collected from Hadza hunter-gatherers in 2013 and 2014; these samples were analyzed in previous publications with different methods (*3*, *5*). A material transfer agreement with the National Institute for Medical Research in Tanzania ensures that the collected stool samples are used solely for academic purposes, and permission for the study was obtained from the National Institute of Medical Research (MR/53i 100/83, NIMR/HQ/R.8a/Vol.IX/1542) and the Tanzania Commission for Science and Technology. Verbal consent was obtained from the Hadza after the study's intent and scope was described with the help of a translator. The publications that first described these samples included several scientists and Tanzanian field guides as coauthors for the critical roles they played in sample collection; however, no new samples were collected in this study and as such, only scientists who contributed to the analyses described here are included as coauthors in this publication. It is currently not possible for us to travel to Tanzania and present our results to the Hadza people; however, we intend to do so once the conditions of the COVID-19 pandemic allow it. **Funding:** This work was supported by the following: National Institutes of Health grants DP1-AT009892 and R01-DK085025 (to J.L.S.); NSF Graduate Research Fellowship grants DGE-1656518 (to D.D.) and DGE-114747 (to B.D.M.); Stanford Graduate Smith Fellowship (to D.D. and M.M.C.); National Institutes of Health grant F32DK128865 (to M.R.O.); Funding was also provided by the Bill and Melinda Gates Foundation. **Author contributions:** Conceptualization: D.D., A.R.J., and J.L.S. Genomic Sequencing: N.N., B.Y., B.D.M., S.T., and D.D. Methodology: D.D., A.R.J., M.R.O., M.M.C., B.D.M., S.J., S.H., and H.W. Data analysis: D.D., M.R.O., M.M.C., B.D.M., S.T., and S.J. Funding Acquisition: D.D., J.L.S., E.D.S., A.R.J., and M.R.O. Supervision: E.D.S., J.L.S., A.R.J., and S.H. Writing - original draft: D.D., M.R.O., E.D.S., and J.L.S. Writing - reviewing and editing: M.R.O., D.D., A.R.J., E.D.S., J.L.S., and M.M.C. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files. Metagenomic reads and genomes generated in this study are available under BioProject PRJEB49206. Accession numbers for individual samples and genomes are available in tables S2 and S3. **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.abj2972
Materials and Methods
Figs. S1 to S10
Tables S1 to S8
References (*37*–*71*)
MDAR Reproducibility Checklist
Data S1 to S8

View/request a protocol for this paper from *Bio-protocol*.

Submitted 3 May 2021; resubmitted 27 January 2022
Accepted 5 May 2022
10.1126/science.abj2972

ORGANIC CHEMISTRY

Doubly stereoconvergent crystallization enabled by asymmetric catalysis

Pedro de Jesús Cruz, William R. Cassels, Chun-Hsing Chen, Jeffrey S. Johnson*

Synthetic methods that enable simultaneous control over multiple stereogenic centers are desirable for the efficient preparation of pharmaceutical compounds. Herein, we report the discovery and development of a catalyst-mediated asymmetric Michael addition/crystallization–induced diastereomer transformation of broad scope. The sequence controls three stereogenic centers, two of which are stereochemically labile. The configurational instability of 1,3-dicarbonyls and nitroalkanes, typically considered a liability in stereoselective synthesis, is productively leveraged by merging enantioselective Brønsted base organocatalysis and thermodynamic stereocontrol using a single convergent crystallization. The synthesis of useful γ-nitro β-keto amides containing three contiguous stereogenic centers is thus achieved from Michael acceptors containing two prochiral centers.

The development of robust, stereoselective synthetic methods that achieve precise control in the simultaneous construction of multiple stereogenic centers is crucial to accelerating the discovery of the next generation of drugs (*1*–*7*). The increase in the stereochemical complexity of such compounds has inspired academic and industrial chemists to invent modern and more efficient methods to facilitate their construction. Enantioselective catalysis is a powerful and broadly applicable paradigm that has been used extensively, and the availability of myriad mechanistic manifolds that can be used in the construction of C–C and C–heteroatom bonds render this blueprint especially attractive (*8*). With notable exceptions (*9*), the discovery of enantioselective catalytic reactions tends to focus on how to maximize stereoselectivity (transition state focus), whereas issues around postreaction processing and optimization of physical properties of the products (ground state focus), considerations that are critical in fields such as polymer chemistry or process chemistry, tend to be neglected. An unfortunate byproduct of this bifurcation in focus is that even for relatively efficient reactions, the practicing chemist is often faced with laborious energy- and resource-intensive purifications that limit applicability on larger scales (*10*). This issue is aggravated when valuable matter is lost as undesired stereoisomers.

To circumvent the inherent issues of ordinary purification techniques (e.g., flash column chromatography, high-pressure liquid chromatography, etc.), the crystallization or precipitation of products from reaction mixtures substantially simplifies isolation while decreasing the amount of waste generated, time spent, and energy required (*11*). For reactions that generate stereoisomeric mixtures, simple recrystallizations cannot overcome the fact that valuable material will be lost as undesired stereoisomers; however, for cases in which equilibration between stereoisomers is mechanistically feasible, convergence to a single stereoisomer of the product can be achieved by engineering crystallization-induced diastereomer transformations (CIDTs). CIDTs are highly desirable in synthetic chemistry and industrial applications because they provide a means to generate highly stereoenriched products without requiring additional tedious purifications (*12*, *13*). CIDT selectivity is governed by crystallization thermodynamics: Diastereomers undergoing CIDT contain one or more static asymmetric centers and at least (and commonly) one labile element of chirality that is subjected to equilibration. Implementation of CIDT strategies into synthetic routes reduces the effort required to access a single stereoisomer of a complex molecule: 100% theoretical yield of a single stereoisomeric product can be obtained from an initial mixture of interconverting epimers. CIDT reactions are usually applied to specific problems in industrial chemistry, they are difficult to predict, and generalizable non-auxiliary–based CIDT methods are currently underdeveloped (*12*, *13*). We were interested in testing the notion that by leveraging epimerization to achieve stereoconvergence in complex systems, we could accrue unique advantages through the merged application of asymmetric catalysis and crystallization-driven selectivity.

The base-catalyzed Michael reaction between two prochiral reaction partners was judged to be an ideal test case to evaluate such a hypothesis: Asymmetric variants comprise atom economical skeletal assemblies in organic synthesis and have been shown to proceed efficiently with a variety of catalytic platforms (*14*). For the projected application, the obligatory electron-withdrawing groups in the starting materials that enable the polar C–C bond construction should also engender multiple acidic C–H sites in the product required to establish the requisite complex equilibria; however, a parallel consideration highlighted in Fig. 1A is that the complexity of the system rises geometrically as the number of configurationally unstable asymmetric centers grows (**I.a** ⇆ **I.b** ⇆ **I.c** ⇆ **P**). For this reason, an enantioselective reaction in which a dual-role catalyst both mediates the installation of a keystone stereocenter and induces completely convergent crystallization of a product with two labile centers is unknown (*15*–*17*). Here, we disclose such an advance, in which the combination of bifunctional Brønsted base asymmetric organocatalysis (*18*) with CIDT principles enables the stereoconvergent syntheses of γ-nitro β-keto amides containing three contiguous asymmetric centers from two prochiral reaction partners through CIDT reactions with considerable scope and downstream utility.

Foundational studies were initiated to assess the viability of a stereoconvergent crystallization using nitromethane (**1a**) as a pronucleophile and the prochiral Michael acceptor **2** (Fig. 1B). Under homogeneous conditions (see the supplementary material for details), the conjugate addition product was obtained in high yields, although the stereochemistry of the β-dicarbonyl stereogenic center was uncontrolled, as expected [85% yield, 1.1:1 diastereomeric ratio (dr)]. The use of ethereal solvents [e.g., diethyl ether, methyl *tert*-butyl ether (MTBE)] allowed for selective crystallization of a single diastereomer of the conjugate addition adduct directly from the reaction medium (see the supplementary materials for solvent studies). Reaction concentrations and temperatures were moreover optimized to prevent the spontaneous precipitation of starting material and isomerically impure product (if the reaction was too concentrated) or loss of product in the filtrate (if the reaction was too dilute). Concurrent with the optimization of the reaction-based crystallization protocol, we also examined the features of the Brønsted base catalyst and their effect on the stereoselectivities and efficiency of the protocol (see the supplementary materials for catalyst optimization). We found that the chiral Dixon iminophosphorane **A** (*18*) successfully engaged in the proposed stereoconvergent crystallization, giving β-keto amide **3a** with excellent yields and enantioselectivity and diastereoselectivity [yield 96%, enantiomeric ratio (er) 94:6, dr >20:1] after a single filtration of the reaction.

This stereoconvergent crystallization protocol could be successfully applied to a range of substituted alkylidenes that were converted into their corresponding crystalline nitro ketone adducts in good to excellent yields and stereoselectivities (Fig. 2). Nitro ketones containing halogens (**3b** and **3k**), alkyl (**3c** and

Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.
*Corresponding author. Email: jsj@unc.edu

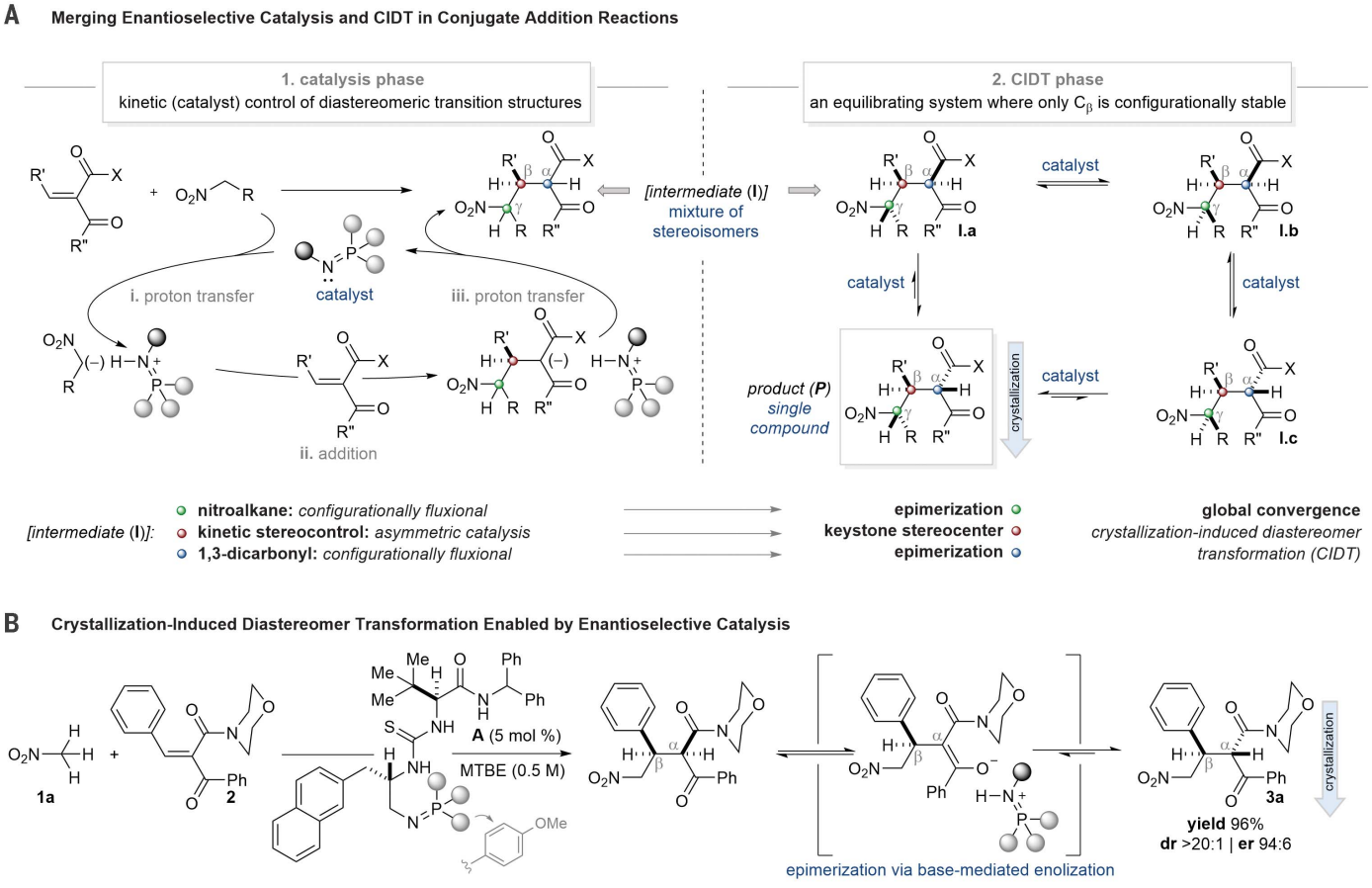


Fig. 1. Proposed mechanism and preliminary studies. (**A**) Combination of Brønsted base enantioselective catalysis and crystallization-induced diastereomer transformations for stereoconvergent reactions. (**B**) Successful combination of asymmetric organocatalysis and CIDT to realize thermodynamically driven stereoconvergence.

3d), and electron-withdrawing (**3e** to **3h**) groups at various positions were obtained in good to excellent yields and stereoselectivities starting from unsymmetrical aryl alkylidenes. A product containing synthetically useful boronic ester **3i** was synthesized in moderate yields and excellent stereoselectivities after the stereoconvergent crystallization procedure. *p*-Dimethylamine– and naphth-2-yl–derived alkylidenes successfully engaged in this reaction delivering nitroalkanes **3j** and **3l** in good yields and stereoselectivities. A variety of heteroaryl products, including furan-3-yl (**3m**), thien-2-yl (**3n**), *N*-methyl pyrrol-2-yl (**3o**), pyrid-2-yl (**3p**), and Boc-protected indol-3-yl (**3q**), were obtained in equally efficient and enantioselective reactions. Several ketone and amide substrates were tested in this reaction (Fig. 3). Aryl ketone products containing halogens (**3r**), alkyl (**3s**), and electron-donating groups (**3t**) were obtained in good to excellent yields and enantioselectivities. Alkyl ketone substrates were tolerated under the reaction conditions, although nitroalkane (**3u**) was obtained at a low enantiomeric ratio. A variety of amides were suitable substrates for this protocol. Piperidine and synthetically useful Weinreb (*19*) amide substrates engaged in fully enantio- and diastereoselective stereoconvergent crystallization to deliver the corresponding nitroalkanes (**3v** and **3w**) in good yields. Last, diisopropyl amide alkylidenes delivered the corresponding nitroalkanes with excellent efficiency (**3x** to **3z**). In some instances (**3y** and **3z**), the diisopropyl amide outperformed the morpholine amide (**3f** and **3k**), and improvements in the enantio- and diastereoselectivity were observed (see below). These preliminary results suggest that the identity of the acyl group could be tuned to improve the efficiency of the crystallization and can be a point of diversification when optimizing related CIDT reactions.

Extensive x-ray diffraction studies were undertaken to evaluate product stereochemistry. As predicted based on a paradigm involving kinetic (catalyst) control of the static asymmetric center, x-ray analysis showed that the β-configuration is rigorously conserved for the Michael addition products **3a**, **3f**, **3h**, **3i**, **3j**, **3n**, **3t**, and **3x**. The fluxional nature of the α-stereocenter could in principle lead to different results across the series based on solubility properties, but the outcomes here were consistent as well: The (*S*)-configuration was observed at the β-keto amide methine in the crystallized products.

Having established a method to control the α-stereogenic center through thermodynamically driven stereoconvergence, attention was then directed to expanding this platform to the use of prochiral nucleophiles. We began our studies with the reaction of *p*-trifluoromethylphenyl alkylidene using nitroethane as the prochiral nucleophile. Under identical conditions (see the supplementary materials), nitroalkane product **4e** crystallized from solution and was obtained in 95% yield, 95:5 er, and 6:1 dr after a single filtration. In situ monitoring revealed that the Michael addition was relatively fast, whereas diastereomerization and crystallization were slower. Accordingly, increasing the catalyst loading to 20 mol % and changing the solvent to 2-methyltetrahydrofuran improved the overall efficiency of the CIDT (see the supplementary materials for additional details), presumably due to accelerated interconversion of the diastereomeric mixture. Under the optimized conditions, the desired conjugate addition product was obtained in excellent yield and stereoselectivity (Fig. 3).

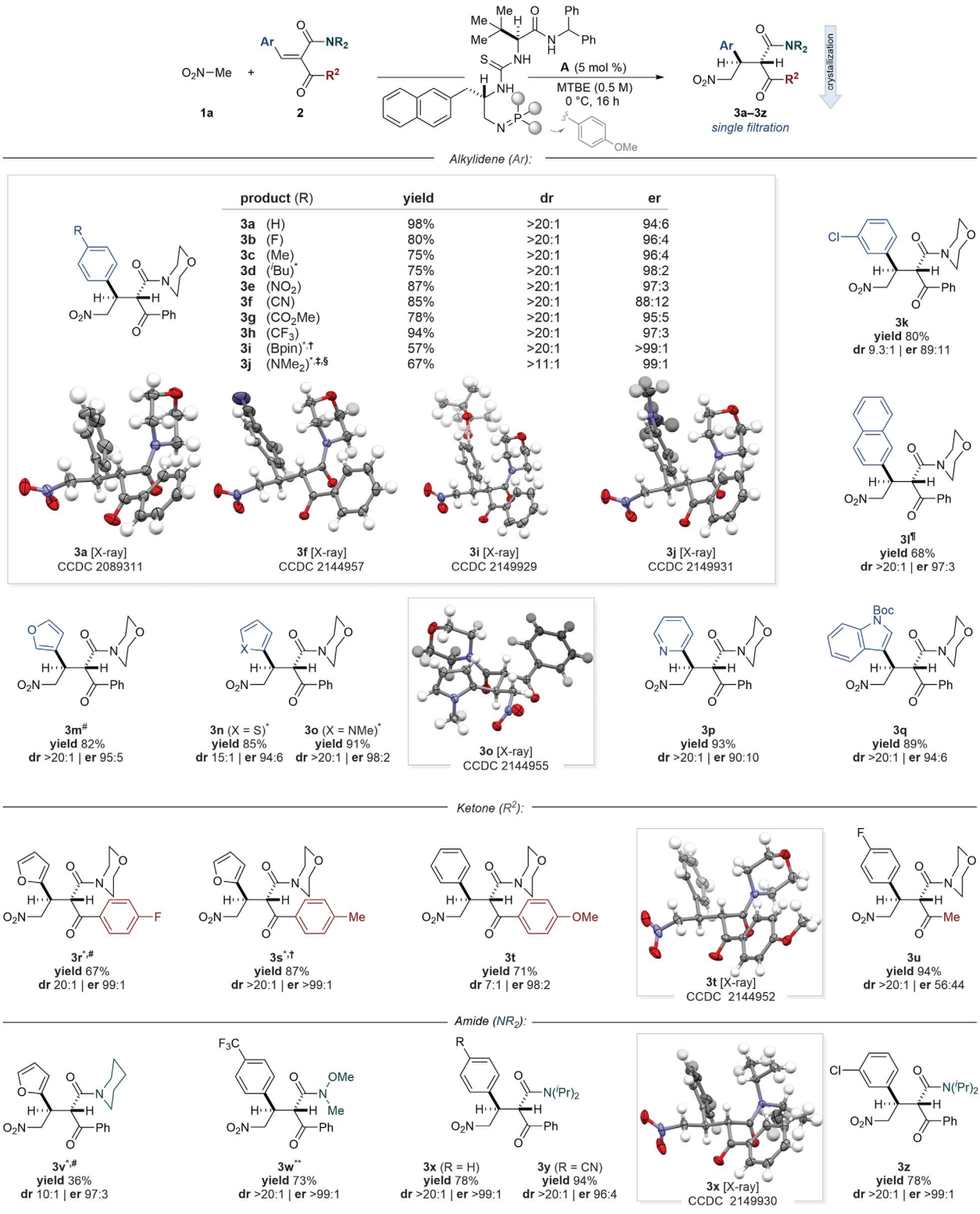


Fig. 2. Substrate scope of the crystallization-enabled stereoconvergent conjugate addition. Reaction conditions were as follows: cat. **A** (5.0 mol %), alkylidene (0.200 mmol, 1.0 equiv, 0.5 M), $MeNO_2$ (0.600 mmol, 3.0 equiv), MTBE (0.400 ml), 0°C, 16 hours. Yields refer to isolated yields. The dr values were determined by ^{1}H NMR spectroscopic analysis of the solid obtained by filtration of the crude reaction mixture. The er values of the solids after filtration were determined by high-performance liquid chromatographic analysis using a chiral stationary phase. *Reaction performed at 23°C. †Reaction performed in 2-Me-THF as solvent. ‡Reaction performed with 5 equiv $MeNO_2$. §Reaction performed using MTBE:DCM (5:1) as a solvent to help solubilize alkylidene starting material. ¶Reaction performed with 15 equiv of $MeNO_2$ to help solubilize alkylidene starting material. #Reaction performed in Et_2O as solvent. **Crystallization occurred after 96 hours.

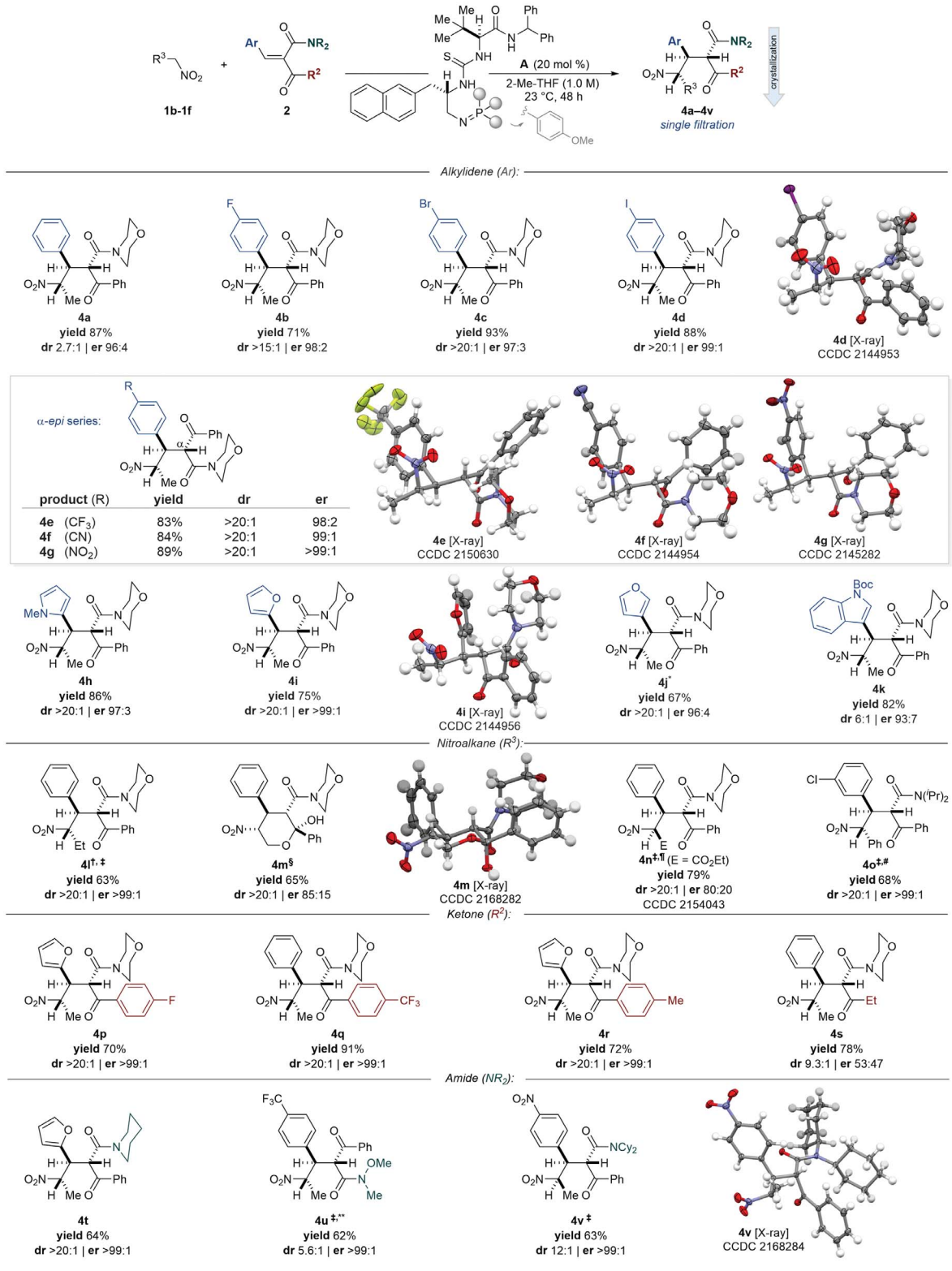


Fig. 3. Substrate scope of the crystallization-enabled doubly stereoconvergent conjugate addition. Reaction conditions were as follows: cat. **A** (20 mol %), alkylidene (0.200 mmol, 1.0 equiv, 1 M), $EtNO_2$ (0.600 mmol, 3.0 equiv), 2-Me-THF (0.200 ml), 23°C, 48 hours. Yields refer to isolated yields. The diastereomeric ratio values were determined by ^{1}H NMR spectroscopic analysis of the solid obtained by filtration of the crude reaction mixture. The er values of the solids after filtration were determined by high-performance liquid chromatographic analysis using a chiral stationary phase. *Reaction performed in Et_2O as solvent. †Reaction performed with 1-nitropropane (0.600 mmol, 3.0 equiv). ‡Reaction performed with MTBE (0.2 ml, 1.0 M). §Reaction performed with nitroethanol (0.600 mmol, 3.0 equiv). ¶Reaction performed with ethyl nitroacetate (0.600 mmol, 3.0 equiv). #Reaction performed with (nitromethyl)benzene (0.600 mmol, 3.0 equiv). **Crystallization observed after 96 hours.

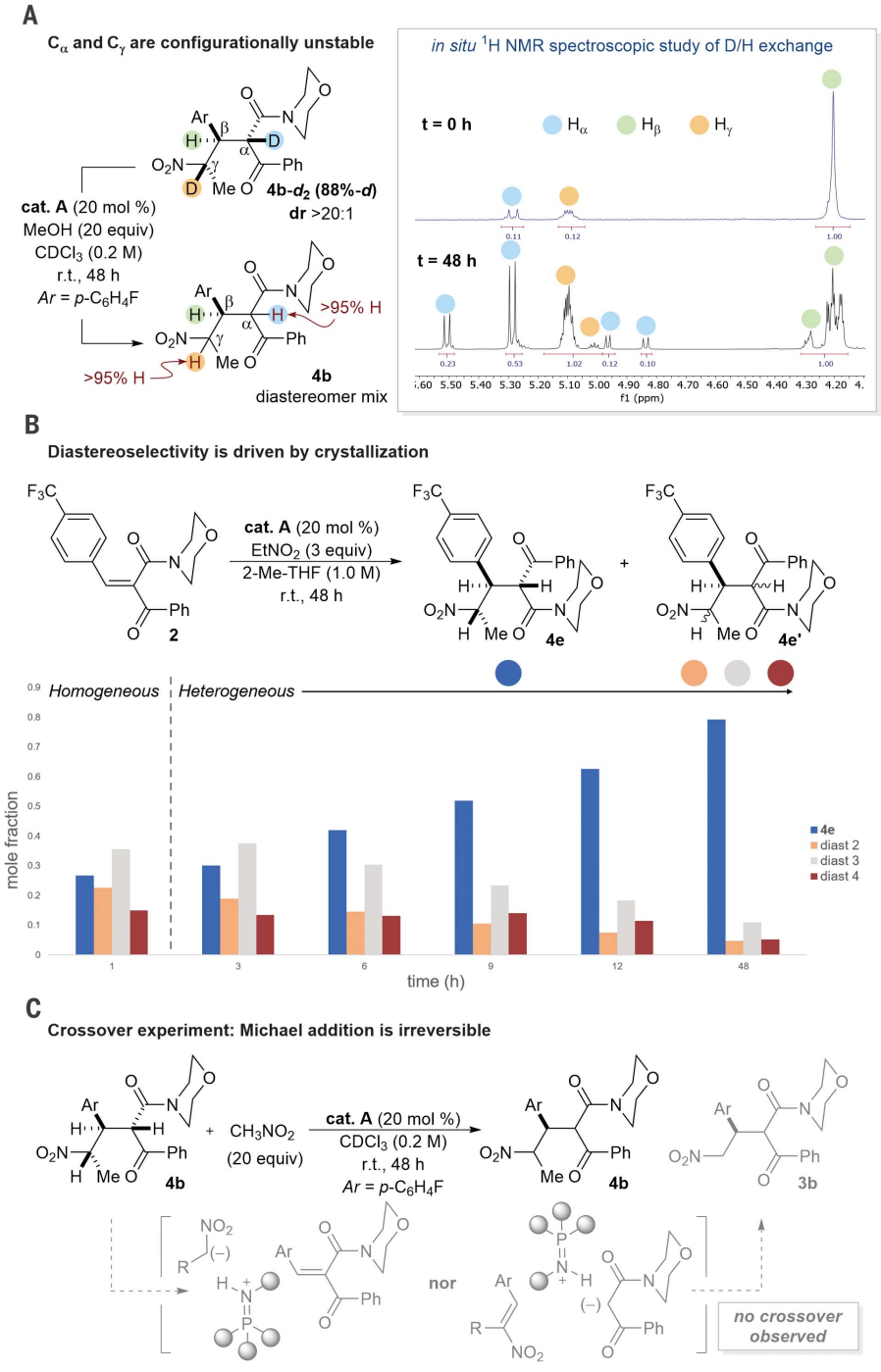


Fig. 4. Mechanistic insights into the origin of stereoselectivity. (**A**) Deuterium/hydrogen exchange studies were performed to establish that the catalyst can epimerize both labile stereocenters. (**B**) An initial diastereomeric mixture converges to a major diastereomer through crystallization-driven stereoconvergence. (**C**) Epimerization through retro-Michael addition is mechanistically possible but not operative.

This result represents a rare example of an efficient noncascade Michael addition reaction between prochiral nucleophilic and electrophilic reaction partners in which the latter is prochiral at both alkene carbon atoms (*14*).

Like the single CIDT reaction, this protocol is applicable to a wide range of substrates (Fig. 3). A variety of aryl alkylidenes delivered nitroalkanes containing halogens (**4b** to **4d**), and electron-withdrawing (**4e** to **4g**) functional groups. Heterocyclic alkylidenes were tolerated (**4h** to **4k**), and substrates including *N*-methyl pyrrole (**4h**), and furan-2-yl (**4i**), furan-3-yl (**4j**), and Boc-protected indol-3-yl (**4k**) were suitable substrates for this transformation. Other prochiral nitroalkanes were tolerated in the doubly stereoconvergent crystallization manifold and delivered β-keto amides containing other alkyl (**4l**), alcohol (**4m**), ester (**4n**), and aryl (**4o**) functional groups in moderate to good yields and stereoselectivities. A variety of aryl ketone substrates bearing halogens (**4p**), electron-withdrawing substitutents (**4q**), and alkyl substituents (**4r**) were also viable substrates for this transformation. Ethyl ketone substrate delivered the desired nitroalkane in excellent yield but low stereoselectivities (**4s**). Finally, a variety of amides were suitable substrates for this transformation, and nitroalkane adducts containing piperidine amide (**4t**), synthetically useful Weinreb amide (**4u**), and dicyclohexyl amide (**4v**) delivered the desired conjugate addition products in good to excellent yield and stereoselectivities.

As in the nitromethane additions, x-ray crystallography was indispensable in assessing the stereochemical outcomes of the high-order CIDTs. Effective regulation of the static asymmetric β-center was again enabled by the Dixon iminophosphorane catalyst. The γ-stereocenter was conserved in the six of the seven analyzed products, with the (*S*)-configuration regularly observed at the nitronate center; the outlier (γ-(*R*) configuration) was nitroalkane **4v**. A comparison of amides **4g** and **4v** revealed inverted configurations at both labile centers in the products, triggered only by changing the amide identity (morpholine amide versus dicyclohexyl amide). Dichotomous stereochemical behavior at the α-center was observed in CIDT reactions giving Michael adducts with electron-poor (**4e** to **4g**) and electron-neutral or -rich aromatic groups (**4d**, **4i**, and **4m**). The assignments for those products not yet studied by x-ray diffraction must be construed as tentative at this point, but the ability to fully invert the obtained major diastereomer in certain cases is exciting, and future work will be directed at understanding and exploiting the structural factors that favor isomer-selective crystallization.

To gain further insight into the proposed doubly stereoconvergent crystallization process, a series of mechanistic experiments was performed (Fig. 4). To understand the stereolability of $\mathbf{C}_\alpha$ and $\mathbf{C}_\gamma$ in the presence of a chiral Brønsted base, deuterium/hydrogen exchange studies using isotopically labeled, diastereomerically pure (dr >20:1) nitroalkane **4b-*d*$_2$** were performed. Using CH_3OH [20 equivalents (equiv)] as the protic additive, >95% H incorporation and stereochemical scrambling at $\mathbf{C}_\alpha$ and $\mathbf{C}_\gamma$ were observed in the presence of the iminophosphorane **A**, indicating that both centers are susceptible to epimerization by the catalyst under the reaction conditions (Fig. 4A). We then turned our attention to understanding the rates of epimerization and the efficiency of the crystallization (Fig. 4B). An identical set of parallel reactions were performed simultaneously and quenched at different time points

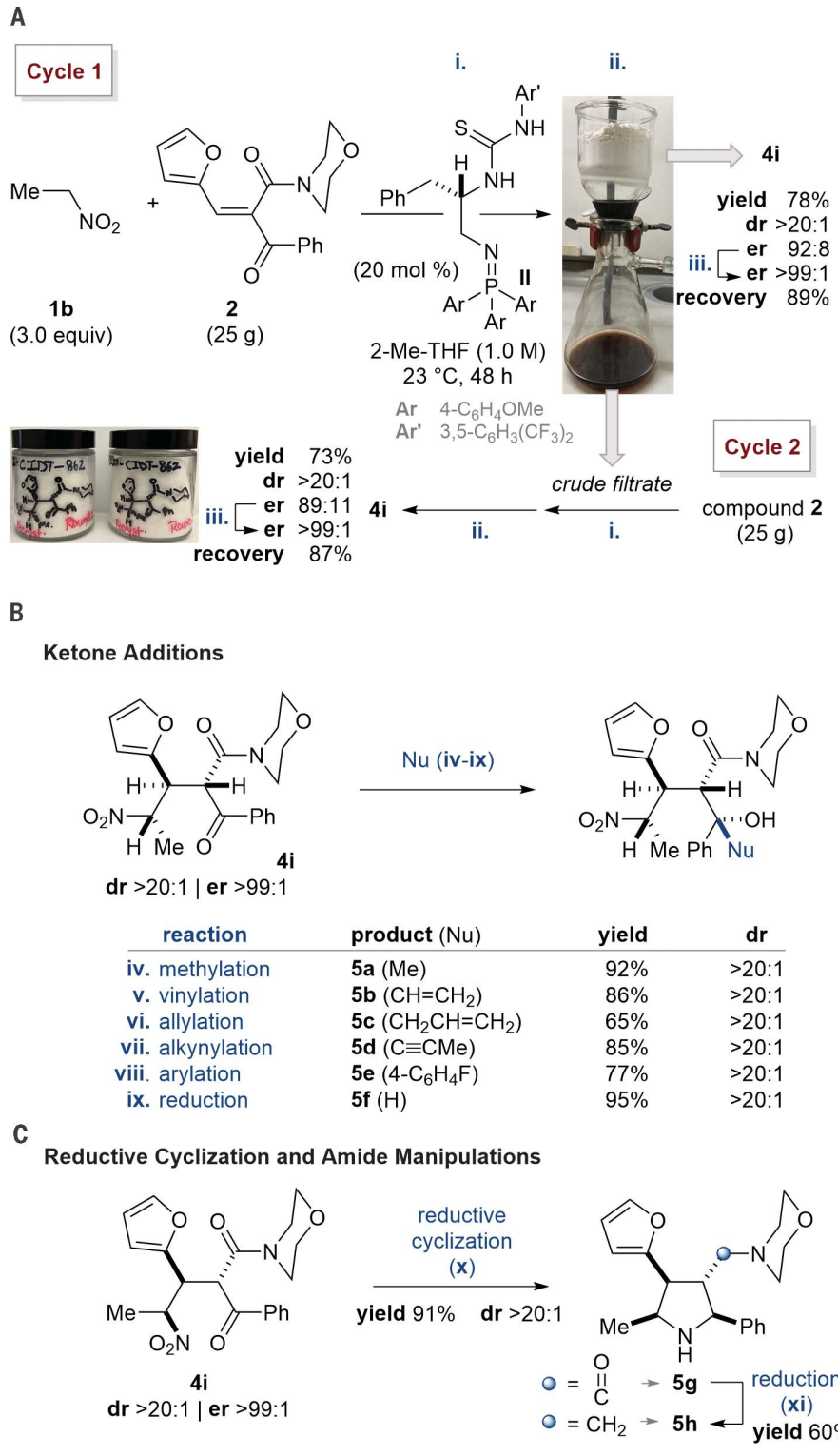


Fig. 5. Synthetic utility of the crystallization-enabled doubly stereoconvergent conjugate addition. See the supplementary materials for specific reaction details. (**A**) A 50 g doubly stereoconvergent crystallization enabled by catalyst recycling. Reaction conditions were as follows: (**i**) cat. **B** (20 mol %), alkylidene **2** (80.3 mmol, 1.0 equiv), $EtNO_2$ (241.0 mmol, 3.0 equiv), 2-Me-THF (80 ml, 1.0 M), 23°C, 48 hours. (**ii**) Unpurified homogeneous filtrate containing catalyst and excess nitroalkane was charged with (80.3 mmol, 1.0 equiv) of prochiral alkylidene **2**. (**iii**) Recrystallization from 80% EtOAc:hexanes. (**B**) Diastereoselective synthesis of secondary and tertiary alcohols. (**iv** to **viii**) nitroalkane **4i** (0.100 mmol, 1.0 equiv), $CeCl_3$ (0.300 mmol, 3.0 equiv), LiCl (0.600 mmol, 6.0 equiv), RMgX (0.180 mmol, 1.8 equiv) THF/DCM (5:1), −78°C, 12 hours. (**ix**) nitroalkane **4i** (0.100 mmol, 1.0 equiv), $TiCl_4$ (0.110 mmol, 1.1 equiv), BH_3•DMS (0.500 mmol, 5.0 equiv), DCM, −78°C, 5 hours. (**C**) Reductive cyclization and synthetic utility of the amide functional handle. (**x**) nitroalkane **4i** (1.5 mmol, 1.0 equiv), Zn (45.0 mmol, 30 equiv), AcOH, 65°C, 5 hours. (**xi**) nitroalkane **4i** (0.200 mmol, 1.0 equiv), $LiAlH_4$ (0.800 mmol, 4.0 equiv), THF, 23°C, 2 hours.

to quantitatively study the product distribution over time using ^{1}H and ^{19}F nuclear magnetic resonance (NMR) spectroscopic techniques. The medium was homogeneous early in the reaction, and no significant preference for any product diastereomer was observed. A mixture of four diastereomers was initially present in solution. The onset of crystallization (t ~ 3 hours) initiated product diastereomer enrichment through CIDT at the expense of the three more soluble, equilibrating diastereomers. Reaction progress was marked by steady perturbation of the isomer ratio away from the unselective solution equilibrium (t = 1 hour), continuing over the course of the 48-hour reaction to result in the highly enriched product.

Epimerization through retro-Michael reversion/Michael addition (*20*) was investigated through a crossover experiment. Upon exposure of nitroethane adduct **4b** to the action of iminophosphorane **A** in the presence of excess nitromethane (20 equiv) under homogeneous reaction conditions for 48 hours, none of the crossover product **3b** was identified; rather, epimerized **4b** and other unidentified decomposition products were observed (Fig. 4C). This experiment underscored an important corollary: Crystallization insulates the product from undesired decomposition pathways that occur in the homogeneous environment commonly favored for organic reactions.

With a mechanistic understanding in hand to rationalize the excellent stereocontrol observed in this reaction platform, an assessment of some of the practical features of the doubly stereoconvergent crystallization was undertaken (Fig. 5). A reaction using 25 g of alkylidene **2** was performed with excess nitroethane. A lower-molecular-weight iminophosphorane (**B**) was used, which efficiently catalyzed the desired CIDT reaction to deliver nitroalkane **4i** in good yield and stereoselectivity after a single filtration. When the unpurified homogeneous filtrate containing catalyst and residual excess nitroethane was treated with another 25 g charge of prochiral alkylidene **2**, nitroalkane **4i** was obtained with nearly identical efficiency in the second cycle. The product was enriched to enantiomeric homogeneity with good recovery

through a single recrystallization (40 g isolated, dr >20:1, er >99:1).

The isolated products of these Michael CIDT reactions can be transformed to value-added products in subsequent steps that capitalize on the embedded functionality without compromising the integrity of the extant asymmetric centers (Fig. 5, B and C) (*21*). Ketone additions using organocerium reagents (*22*, *23*) delivered a variety of tertiary alcohols containing synthetically useful functional handles, including the alkyl (**5a**), vinyl (**5b**), allyl (**5c**), alkynyl (**5d**), and aryl (**5e**) groups, in good to excellent diastereoselectivity. Secondary β-hydroxy amide **5f** was obtained in 95% yield as a single diastereomer through titanium (IV)-mediated diastereoselective reduction (*24*). Pyrrolidine **5g** was synthesized in a single step in excellent diastereoselectivity through Zn-mediated reductive cyclization (*25*), and from that point the amide could be converted to its derived tertiary amine **5h** in moderate yield.

This work establishes a foundation for crystallization-induced diastereomer transformations operating on two configurationally labile asymmetric centers, enabled in this instance by the Dixon chiral iminophosphorane Brønsted superbase. The results of the present study suggest that expanded opportunities may exist for the productive merger of divergent, partially selective first-stage asymmetric catalysis with crystallization-driven second-stage stereoconvergence. A key to the generalization and future growth of such platforms that capitalize on their myriad benefits will be the development of robust predictive tools based on, among other things, analysis of crystal packing and machine learning.

REFERENCES AND NOTES

1. V. Farina, J. T. Reeves, C. H. Senanayake, J. J. Song, *Chem. Rev.* **106**, 2734–2793 (2006).
2. J. Blacker, C. E. Headley, in *Green Chemistry in the Pharmaceutical Industry*, A. S. Wells, M. T. Williams, Eds. (Wiley, 2010); pp. 269–288.
3. M. D. Eastgate, M. A. Schmidt, K. R. Fandrick, *Nat. Rev. Chem.* **1**, 0016 (2017).
4. O. Méndez-Lucio, J. L. Medina-Franco, *Drug Discov. Today* **22**, 120–126 (2017).
5. S. Caille *et al*., *J. Org. Chem.* **84**, 4583–4603 (2019).
6. S. Caille, *Science* **364**, 635 (2019).
7. K. Lovato, P. S. Fier, K. M. Maloney, *Nat. Rev. Chem.* **5**, 546–563 (2021).
8. P. J. Walsh, M. C. Kozlowski, *Fundamentals of Asymmetric Catalysis* (University Science Books, 2009).
9. K. B. Hansen *et al*., *J. Am. Chem. Soc.* **131**, 8798–8804 (2009).
10. G. P. Howell, *Org. Process Res. Dev.* **16**, 1258–1272 (2012).
11. D. J. Kirwan, C. J. Orella, in *Handbook of Industrial Crystallization*, A. S. Myerson, Ed. (Elsevier, ed. 2, 2002); pp. 249–266.
12. A. Kolarovič, P. Jakubec, *Adv. Synth. Catal.* **363**, 4110–4158 (2021).
13. K. M. J. Brands, A. J. Davies, *Chem. Rev.* **106**, 2711–2733 (2006).
14. E. Reyes, U. Uria, J. L. Vicario, L. Carrillo, *Org. React.* **90**, 1–898 (2016).
15. H. T. Openshaw, N. Whittaker, *J. Chem. Soc.* 1461–1471 (1963).
16. P. J. Pye *et al*., *Chemistry* **8**, 1372–1376 (2002).
17. M. C. Prieto-Ramírez *et al*., *Adv. Synth. Catal.* **360**, 4362–4371 (2018).
18. M. Formica, D. Rozsar, G. Su, A. J. M. Farley, D. J. Dixon, *Acc. Chem. Res.* **53**, 2235–2247 (2020).
19. S. Nahm, S. M. Weinreb, *Tetrahedron Lett.* **22**, 3815–3818 (1981).
20. S. P. Lathrop, T. Rovis, *J. Am. Chem. Soc.* **131**, 13628–13630 (2009).
21. D. A. Evans, M. D. Ennis, T. Le, N. Mandel, G. Mandel, *J. Am. Chem. Soc.* **106**, 1154–1156 (1984).
22. T. Imamoto *et al*., *J. Org. Chem.* **49**, 3904–3912 (1984).
23. T. Imamoto, Y. Sugiura, N. Takiyama, *Tetrahedron Lett.* **25**, 4233–4236 (1984).
24. G. Bartoli *et al*., *Tetrahedron Lett.* **42**, 8811–8815 (2001).
25. Z. X. Jia *et al*., *Chemistry* **18**, 12958–12961 (2012).

ACKNOWLEDGMENTS

We thank B. Ehrmann and D. Weatherspoon (UNC Department of Chemistry Mass Spectrometry Core Laboratory) for assistance with mass spectrometry analysis; M. ter Horst (UNC Department of Chemistry NMR Core Laboratory) for assistance with NMR analysis; and H. King (UNC Department of Chemistry) for experimental contributions. **Funding:** This work was supported by the National Institute of General Medical Sciences, National Institutes of Health (grant R35 GM 118055 to J.S.J. and grant F31 GM137697 to P.d.J.C.). **Author contributions:** P.d.J.C. and J.S.J. conceived the study. P.d.J.C., W.R.C., and J.S.J. designed, implemented, and analyzed the experiments. C.H.C. performed the x-ray diffraction studies. P.d.J.C., W.R.C., and J.S.J. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Experimental procedures and characterization data are available in the supplementary materials. The crystallographic data for this paper can be obtained free of charge from the Cambridge Crystallographic Data Centre (www.ccdc.cam.ac.uk/data_request/cif) using the following accession codes: CCDC 2089311, 2128977, 2144952-2144957, 2145282, 2149929-2149931, 2154042-2154043, 2168282, and 2168284. **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.abo5048
Materials and Methods
Figs. S1 to S8
Tables S1 to S19
NMR Spectra
References (*26*–*50*)

Submitted 7 February 2022; accepted 6 May 2022
10.1126/science.abo5048

PHOSPHORUS CHEMISTRY

Enantioselective hydrogen-bond-donor catalysis to access diverse stereogenic-at-P(V) compounds

Katherine C. Forbes and Eric N. Jacobsen*

The stereoselective synthesis of molecules bearing stereogenic phosphorus(V) centers represents an enduring challenge in organic chemistry. Although stereospecific nucleophilic substitution at P(V) provides a general strategy for elaborating optically active P(V) compounds, existing methods for accessing the requisite chiral building blocks rely almost entirely on diastereocontrol using chiral auxiliaries. Catalytic, enantioselective methods for the synthesis of synthetically versatile stereogenic P(V) building blocks offer an alternative approach to stereogenic-at-P(V) targets without requiring stoichiometric quantities of chiral control elements. Here, we report an enantioselective hydrogen-bond-donor–catalyzed synthesis of aryl chlorophosphonamidates and the development of these products as versatile chiral P(V) building blocks. We demonstrate that the two leaving groups on these chlorophosphonamidates can be displaced sequentially and stereospecifically to access a wide variety of stereogenic-at-P(V) compounds featuring diverse substitution patterns.

Phosphorus(V) stereocenters are present in a wide assortment of important molecules, including several recently developed pharmaceuticals (Fig. 1A). The absolute stereochemistry at phosphorus is often directly associated with the biological activity of those molecules (*1*–*7*). Stereogenic-at-phosphorus compounds also serve as broadly useful ligands and catalysts in asymmetric organic synthesis (*8*, *9*). Although a variety of natural products bearing *P*-stereogenic centers have been identified (*10*), these molecules are not practical synthetic building blocks owing to their sparsity. Thus, whereas the synthesis of compounds bearing *C*-stereogenic centers has historically drawn heavily on nature's chiral pool (*11*), access to *P*-stereogenic molecules relies entirely on de novo synthesis. Nucleophilic substitution at stereogenic P(V) centers can occur stereospecifically, thereby providing a powerful strategy for the synthesis of complex, optically active compounds from simple P(V) building blocks bearing one or more leaving groups attached to phosphorus (*9*, *11*–*13*).

Effective methods for accessing stereogenic-at-phosphorus targets have relied primarily on the use of covalently attached chiral auxiliaries to achieve diastereocontrol, and a variety of chelating auxiliaries have been developed successfully for this purpose (Fig. 1B) (*14*–*22*). Their applicability depends on stereospecific displacement of the auxiliary to forge P(V) stereocenters with absolute stereocontrol. Among recent advances using the chiral auxiliary approach, Baran and colleagues reported the development of highly reactive oxathiaphospholane-sulfide building blocks (*19*, *20*). The propensity of the P–S bonds in these building blocks to undergo substitution by both alcohols and organometallic reagents was demonstrated and enables the synthesis of a variety of stereogenic-at-P(V) compounds, ranging from oligonucleotides to chiral phosphine oxides.

Despite important advances in the stereoselective synthesis of chiral P(V) compounds by the chiral auxiliary approach, there are both practical and fundamental motivations for developing asymmetric catalytic strategies toward these targets. In that vein, there have been several recent breakthroughs (Fig. 1B). DiRocco and co-workers developed a chiral bisimidazole–catalyzed synthesis of phosphoramidate prodrugs through the diastereoselective addition of nucleosides to chlorophosphoramidates, proceeding via a cooperative mechanism of covalent activation of P(V) and general-base activation of the alcohol nucleophile (*23*). An alternative approach was demonstrated by Miller and co-workers in the catalytic, stereodivergent synthesis of *P*-stereogenic oligonucleotides from phosphoramidites via chiral phosphoric acid catalysis (*24*). Finally, in work that appeared as the present study was being completed, Dixon and co-workers reported a catalytic, enantioselective desymmetrization of diaryl phosphonate esters by substitution with *ortho*-substituted phenols (*25*). Although high levels of stereoselectivity were achieved in these catalytic, nucleophilic substitution reactions, each is limited to a narrow class of nucleophiles that are not further displaced. We conceived that the catalytic, enantioselective installation of a nucleophile that could further serve as a leaving group for stereospecific substitution at P(V) could provide a generalizable strategy for the synthesis of chiral P(V) targets with the broad synthetic scope of state-of-the-art auxiliary approaches while avoiding the need for the stoichiometric use of chiral control elements.

We selected chlorophosphonamidates as potential targets of an enantioselective catalytic approach (Fig. 1C). The chloride and amino groups on P(V) display orthogonal reactivity that might permit sequential and stereospecific displacement en route to chiral P(V) targets bearing a broad range of substitution patterns. Given that P–Cl bonds in particular are susceptible to substitution by a wide variety of nucleophiles (*26*–*28*), chlorophosphonamidates would be highly versatile precursors to a multitude of P(V) frameworks. We report here the development of an enantioselective method for the synthesis of chlorophosphonamidate intermediates using a commercially available hydrogen-bond (H-bond) donor catalyst, as well as the application of these P(V) building blocks to the synthesis of P(V) compounds featuring diverse substitution patterns.

We recognized that a most concise enantioselective synthesis of chlorophosphonamidates would be realized using a catalytic desymmetrization reaction of phosphonic dichlorides with amines. Dual H-bond donor catalysts have been applied broadly and successfully to promote stereoselective nucleophilic substitution reactions via chloride-abstraction pathways (*29*–*32*), and we hypothesized that this reactivity principle could serve to activate one of the two enantiotopic chlorides of a phosphonic dichloride electrophile toward displacement by an amine. Phenyl phosphonic dichloride **2a** was selected as a model substrate in reactions with various amine nucleophiles and potential chiral catalysts (Fig. 2). The chlorophosphonamidate products were found to be too reactive to isolate in pure form, but solutions of ***3*** were stable and could be separated from other reaction components by filtration through silica. Epimerization of chlorophosphonamidate ***3*** was not observed under the catalytic conditions, even in the presence of added tetrabutylammonium chloride. However, concentrated solutions of ***3*** underwent racemization slowly at room temperature over several hours (table S10). For purposes of isolation and analysis, the chlorophosphonamidates were quenched with sodium methoxide at low temperature to produce the corresponding phosphonamidates (e.g., **4a**). After systematic evaluation of a series of chiral dual H-bond donor catalysts and amine nucleophiles, the sulfinamido urea **1a** (*33*, *34*) was found to promote the nucleophilic substitution by diisoamylamine in 95% enantiomeric excess (ee) and quantitative yield (Fig. 2A; see supplementary materials for optimization studies). Multiple equivalents of amine were required to attain full conversion of **2a**, as the amine functions both as a nucleophile and as a stoichiometric Brønsted base to trap the HCl by-product produced in the reaction. Examination of the role of catalyst structure revealed the importance of both

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA.
*Corresponding author. Email: jacobsen@chemistry.harvard.edu

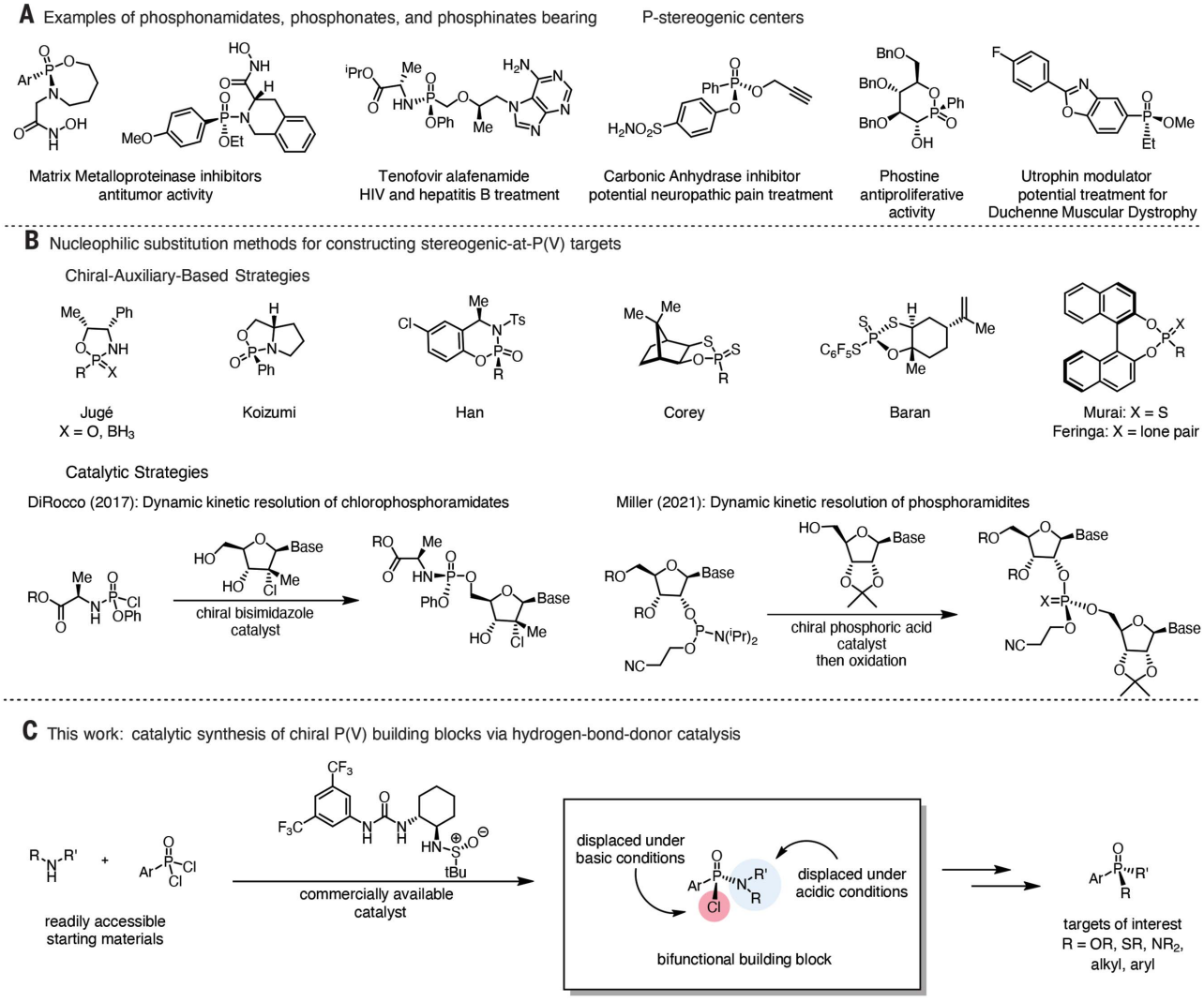


Fig. 1. Methods for accessing stereogenic P(V) targets. (**A**) Representative bioactive compounds bearing *P*-stereogenic centers. (**B**) Synthetic approaches to stereogenic-at-P(V) targets using chiral auxiliaries (*14*–*22*) and stereoselective catalysis (*23*, *24*). (**C**) A general approach to chiral P(V) building blocks via enantioselective catalysis or stereospecific substitution. Ar, aryl; OH, hydroxy group; MeO, methoxy group; OEt, ethoxy group; ^{i}PrO, isopropoxy group; Me, methyl; OPh, phenoxy group; Ph, phenyl; BnO, benzyloxy group; Et, ethyl; OMe, methoxy group; R, alkyl group; Ts, *para*-tolylsulfonyl; RO, alkoxy group; *t*Bu, *tert*-butyl.

the H-bond donor and the sulfinamide group in promoting high enantioselectivity. Whereas sulfinamido urea **1a** and its thiourea analog **1b** proved similarly effective as catalysts, the sulfinamide **1d** lacking the H-bond donor motif induced little acceleration above the uncatalyzed rate (83 versus 64% yield after 24 hours) and afforded only racemic product. The sulfinamido urea **1c** epimeric to **1a** also induced severely diminished enantioselectivity, a stereochemical "mismatch" effect that has also been observed in other applications of this catalyst (*33*, *34*) and one that is strongly suggestive of cooperative participation of the H-bond donor and the sulfinamide in the enantiodetermining step. Arylpyrrolidino (thio)ureas such as **1e**, **1f**, and **1g**, which have proven useful in a wide range of asymmetric anion-binding pathways (*35*) but lack the sulfinamide moiety, were catalytically active but generally poorly effective with respect to enantiocontrol. The enantioselectivity of the substitution was also closely tied to the identity of the amine, with diisoamylamine undergoing reaction with distinctly superior results relative to any of the other nucleophiles examined (Fig. 2B). Beyond a beneficial effect of distal alkyl branching, it is difficult to discern any straightforward correlation between the steric or electronic properties of the amine and enantioselectivity in the substitution reaction. Control studies suggest that the properties of the dialkylammonium chloride by-products likely play a critical and complex role in influencing the observed enantioselectivity, either as inhibitors of the anion-binding H-bond donor catalyst or by promoting a racemic pathway between **2a** and the dialkylamine (tables S8 and S9).

High levels of enantioselectivity were achieved in the reaction of a variety of aryl phosphonic dichlorides with diisoamylamine (Fig. 3A). Substrates bearing arenes with either electron-withdrawing or electron-donating substituents underwent substitution with consistently high levels of enantioselectivity (**4b** to **4g**). In contrast, alkyl phosphonic dichlorides are ineffective substrates for the enantioselective reaction. For example, hexylphosphonic dichloride was converted to the corresponding phosphonamidate in only 26% ee and 50% yield under the catalytic conditions.

The products of the enantioselective reactions feature two chemically distinct leaving

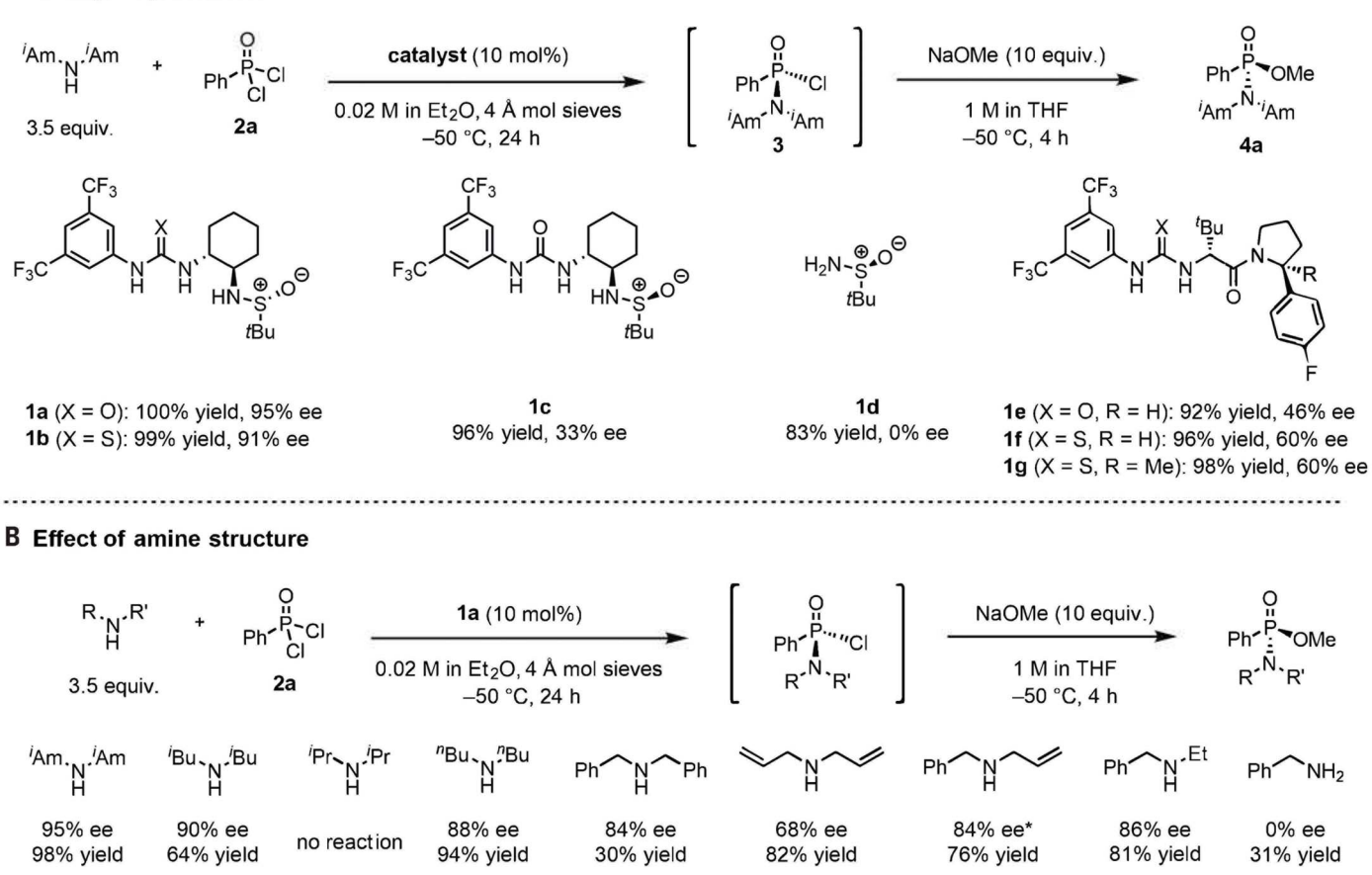


Fig. 2. Optimization studies. Yield values reflect product quantification by ^{31}P nuclear magnetic resonance relative to an internal standard. Reactions were carried out using a one-pot procedure without purification of **3**. Concentration values correspond to the initial concentration of the limiting stoichiometric reagent. (**A**) Catalyst optimization for enantioselective reaction of diisoamylamine with phenyl phosphonic dichloride. Reactions were carried out on a 0.06-mmol scale. (**B**) Optimization of amine structure for enantioselective substitution reaction with phenyl phosphonic dichloride. Reactions were carried out on a 0.06-mmol scale. The single-asterisk symbol indicates that reaction was performed at −40°C for 48 hours. R′, alkyl group; ^{i}Am, isoamyl; Et_2O, diethyl ether; THF, tetrahydrofuran; ^{i}Bu, isobutyl; ^{i}Pr, isopropyl. ^{n}Bu, *n*-butyl.

groups on phosphorus that could be selectively and stereospecifically displaced to afford access to multiple classes of chiral P(V) compounds. We first explored the scope of nucleophiles capable of enantiospecific displacement of the remaining chloride (Fig. 3B). Reaction of ***3*** with alkoxides, phenoxides, thiolates, deprotonated carbamates, and Grignard reagents afforded the desired products with high levels of enantiospecificity (es) in all cases (**5a** to **5h**). The substitution reactions could be performed after the enantioselective catalytic step with or without purification of ***3*** in solution (see supplementary materials for details). We found that the reactions could be scaled up without loss of enantioselectivity or yield; thus, the synthesis of **5d** was performed by the one-pot procedure on a 3-mmol scale with 5 mol % catalyst, affording 1.11 g of product in 95% yield and 92% ee (Fig. 3D).

The products of the chloride-displacement reactions could be further elaborated to afford alkoxy-substituted P(V) compounds via an acid-mediated stereoinvertive displacement of the diisoamylamino group (Fig. 3C). Substitution of **5a** to **5h** with methanol yielded a variety of enantioenriched phosphonates, phosphinates, and phosphonamidates (**6a** to **6h**) with nearly complete enantiospecificity observed in every case. The slightly diminished stereospecificity observed with **5g** and **5h** is consistent with prior observations (*14*, *16*). Substitution with other primary alcohols proceeded with varied but generally high levels of enantiospecificity (**6i** to **6k**).

The phosphonate ester and thioester products **6b** and **6d** have additional readily displaceable substituents that render them useful synthetic building blocks for further elaboration to chiral P(V) compounds. For example, phosphonate thioester **6d** underwent reaction with functionally complex alcohols to furnish the corresponding phosphonylated biomolecules with high levels of stereospecificity (**7a** to **7c**) (Fig. 4A). These substitutions are performed under Brønsted acid–free conditions using little or no excess of the alcohol reagent, highlighting the utility of **6d** for the phosphonylation of precious or acid-sensitive alcohols. Phosphonate **6b** underwent efficient substitution with Grignard reagents with displacement of the electron-deficient aryloxide to yield highly enantioenriched phosphinate esters, known precursors to chiral phosphine oxides (Fig. 4B) (*20*). This three-step route to phosphinate esters was applied to the synthesis of (+)-SMT022332, a utrophin modulator developed as a potential treatment for Duchenne muscular dystrophy (*36*–*38*). An analog of (+)-SMT022332 was previously accessed in 83% ee and 5% overall yield from **9** using a chiral auxiliary–based approach (*4*). Subjection of phosphonic dichloride **9** to the optimized conditions for the enantioselective substitution yielded phosphonamidate **10**, which was characterized crystallographically (Fig. 4C). Subsequent methanolysis and phenol displacement furnished (+)-SMT022332 (**12**) in 94% ee and 43% overall yield over three steps.

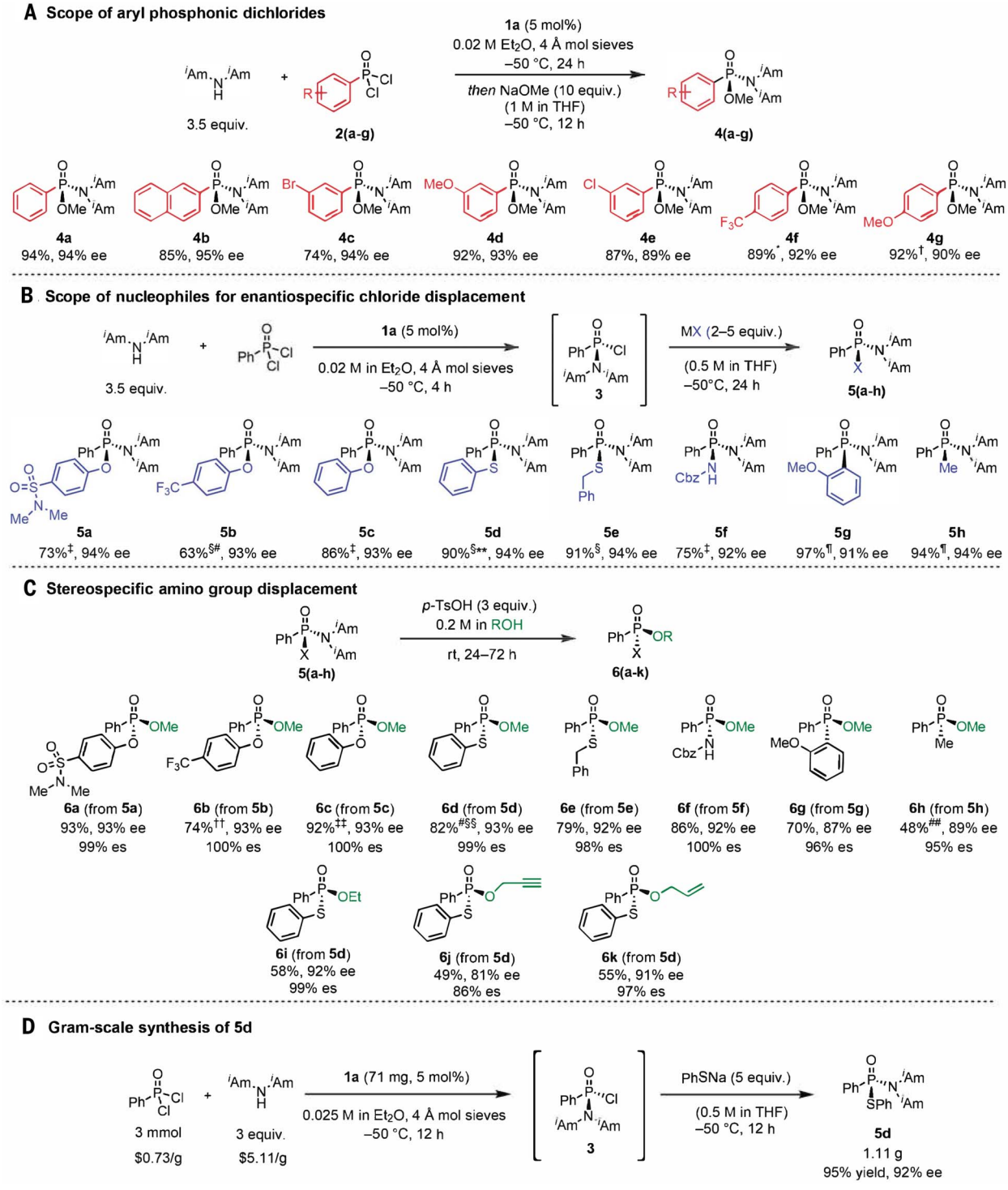


Fig. 3. Scope of enantioselective addition of diisoamylamine to aryl phosphonic dichlorides and stereospecific elaborations. All yield values correspond to chromatographically purified, isolated products. Concentration values correspond to the initial concentration of the limiting stoichiometric reagent. (**A**) Substrate scope of addition of diisoamylamine to aryl phosphonic dichlorides catalyzed by **1a**. Reactions were carried out on a 0.2-mmol scale. The absolute stereochemistry of the products was assigned on the basis of the x-ray crystal structure of **10** and the known optical rotation of **8a** (Fig. 4; see supplementary materials). (**B**) Scope of nucleophiles for enantiospecific substitution with **3**. (**C**) Enantiospecific displacement of the diisoamylamino group with alcohols. See supplementary materials for reaction conditions. rt, room temperature. (**D**) Gram-scale synthesis of **5d**. Prices from Thermo Fisher Scientific (February 2022). The symbols in the figure indicate reaction carried out under the following conditions: *, at −78°C with 20 mol % catalyst loading; †, at −40°C with 4.5 equiv of diisoamylamine; ‡, in a two-pot procedure involving generation of **3** in solution and purification by filtration through silica and subsequent reaction with 2 equiv of nucleophile; §, using one-pot procedure without purification of **3** with 5 equiv of nucleophile; ¶, in a two-pot procedure involving generation of **3** in solution and purification by filtration through silica and subsequent reaction with 5 equiv of nucleophile; #, on a 0.9-mmol scale; **, on a 1.0-mmol scale; ††, on a 0.57-mmol scale; ‡‡, on 0.24-mmol scale; §§, run at 0.3 M concentration instead of 0.2 M; ##, H_3PO_3 was used instead of *para*-tolylsulfonic acid.

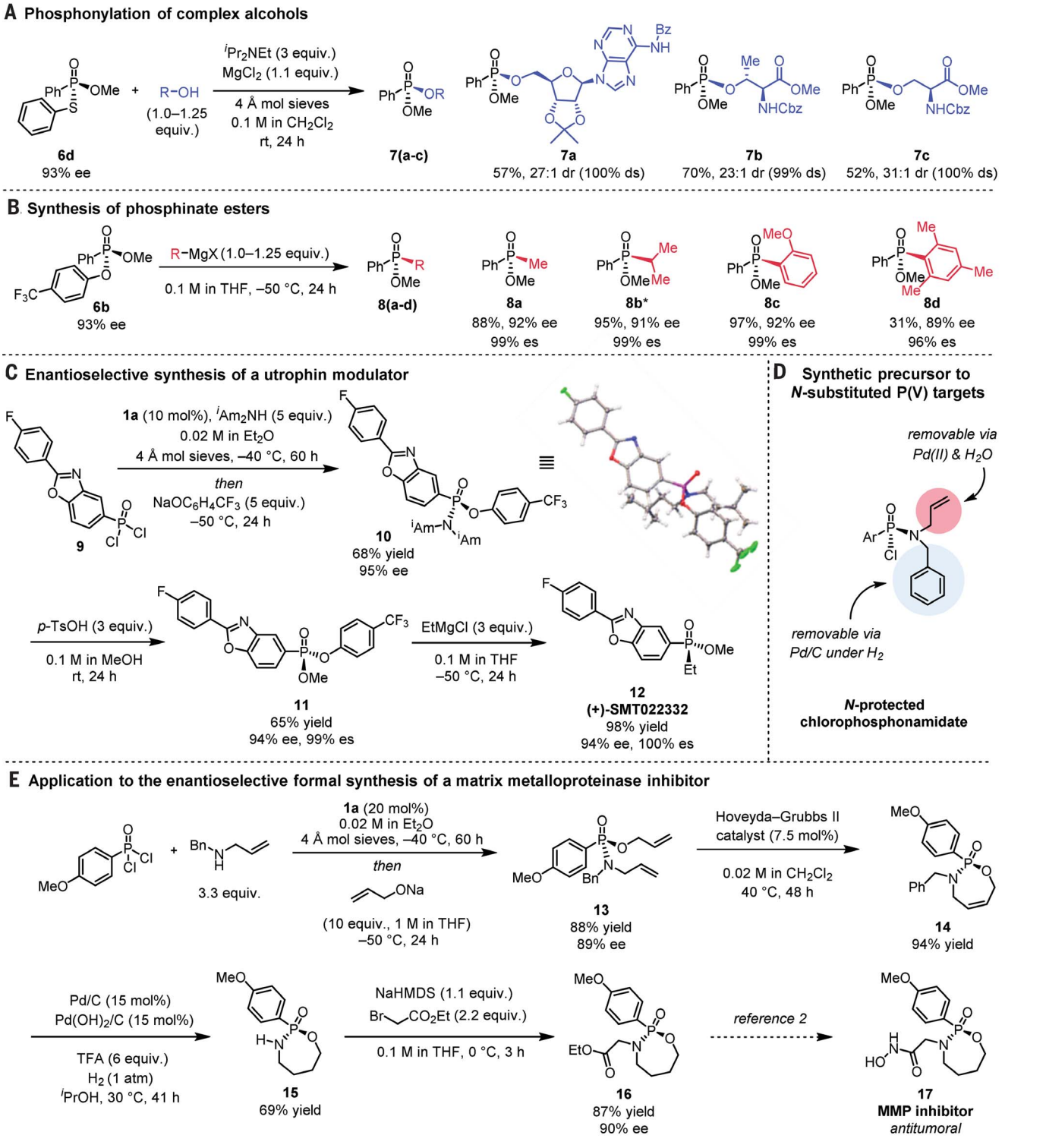


Fig. 4. Application to the synthesis of chiral P(V) targets. All yield values refer to chromatographically purified, isolated products. Concentration values correspond to the initial concentration of the limiting stoichiometric reagent. (**A**) Stereospecific phosphonylation of precious alcohols with **6d**. Reactions were carried out on a 0.1-mmol scale. (**B**) Stereospecific addition of Grignard reagents to **6b** for the synthesis of enantioenriched phosphinate esters. Absolute stereochemistry of **8a** was determined by comparison of optical rotation to literature value; others were assigned by analogy. Reactions run on a 0.05- to 0.1-mmol scale. The single-asterisk symbol indicates product prepared from **6b** that was 92% ee. (**C**) Application of method to the enantioselective synthesis of (+)-SMT022332. Yield values refer to isolated yields. Absolute stereochemistry of **10** assigned by the depicted x-ray crystal structure, and of **12** by comparison of the optical rotation to the literature value. (**D**) Orthogonally *N*-protected chlorophosphonamidate. (**E**) Formal synthesis of a matrix metalloproteinase inhibitor. dr, diastereomeric ratio; ds, diastereospecificity; TFA, trifluoroacetic acid.

In addition to serving as versatile synthetic building blocks, phosphonamidates are often synthetic targets themselves (*2*, *3*, *5*, *6*, *39*–*43*), and general access to these compounds by the catalytic procedure would be desirable. However, the structural requirements on the amine for achieving high enantioselectivity in the catalytic reaction impose restrictions to the *N*-substituents that can be introduced directly (Fig. 2B). We therefore sought to identify amine derivatives that participate successfully in the enantioselective reaction while bearing orthogonally cleavable *N*-protecting groups that might provide centralized access to a variety of substituted phosphonamidates (Fig. 4D). High enantioselectivity was obtained using *N*-allyl benzylamine in the substitution reaction under modified conditions. The benzyl group and the allyl group on the chlorophosphonamidate products can each be cleaved successively, enabling their sequential replacement (see supplementary materials) (*44*–*48*). This strategy was exploited in the synthesis of phosphonamidate **17**, a matrix metalloproteinase (MMP) inhibitor with demonstrated anticancer activity (Fig. 4E) (*2*). Phosphonic dichloride **2h** effectively underwent the catalytic reaction with *N*-allyl benzylamine to produce, after quenching with allyl alkoxide, phosphonamidate **13** in 89% ee and 88% yield. Phosphonamidate **13** was elaborated over three steps to afford cyclic phosphonamidate **16** in 90% ee, completing the enantioselective formal synthesis of MMP inhibitor **17**. We anticipate that *N*-allyl benzylamine's versatility as a masked "–NH_2" equivalent may enable access to a wide variety of phosphonamidate targets.

Moreover, we expect the versatile enantioenriched chlorophosphonamidate intermediates accessed by means of the synthetic strategies outlined herein to enable the facile synthesis of both known and new stereogenic-at-P(V) compounds of interest.

REFERENCES AND NOTES

1. U. Pradere, E. C. Garnier-Amblard, S. J. Coats, F. Amblard, R. F. Schinazi, *Chem. Rev.* **114**, 9154–9218 (2014).
2. M. D. Sørensen *et al*., *Bioorg. Med. Chem.* **11**, 5461–5484 (2003).
3. M. Sawa *et al*., *J. Med. Chem.* **45**, 919–929 (2002).
4. A. Babbs *et al*., *Tetrahedron* **76**, 130819 (2020).
5. A. Nocentini *et al*., *J. Med. Chem.* **63**, 5185–5200 (2020).
6. A. Nocentini, P. Gratteri, C. T. Supuran, *Chemistry* **25**, 1188–1192 (2019).
7. W. A. Lee *et al*., *Antimicrob. Agents Chemother.* **49**, 1898–1906 (2005).
8. T. Imamoto, *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **97**, 520–542 (2021).
9. M. Dutartre, J. Bayardon, S. Jugé, *Chem. Soc. Rev.* **45**, 5771–5794 (2016).
10. O. I. Kolodiazhnyi, *Symmetry* **13**, 889 (2021). 10.3390/sym13050889
11. O. I. Kolodiazhnyi, A. Kolodiazhna, *Tetrahedron Asymmetry* **28**, 1651–1674 (2017).
12. O. I. Kolodiazhnyi, *Tetrahedron Asymmetry* **23**, 1–46 (2012).
13. X. Ye, L. Peng, X. Bao, C.-H. Tan, H. Wang, *Green Synth. Catal.* **2**, 6–18 (2021).
14. S. Jugé, J. P. Genet, *Tetrahedron Lett.* **30**, 2783–2786 (1989).
15. S. Jugé, M. Stephan, J. A. Laffitte, J. P. Genet, *Tetrahedron Lett.* **31**, 6357–6360 (1990).
16. T. Koizumi, R. Yanada(nee Ishizaka), H. Takagi, H. Hirai, E. Yoshii, *Tetrahedron Lett.* **22**, 571–572 (1981).
17. Z. S. Han *et al*., *J. Am. Chem. Soc.* **135**, 2474–2477 (2013).
18. E. J. Corey, Z. Chen, G. J. Tanoury, *J. Am. Chem. Soc.* **115**, 11000–11001 (1993).
19. K. W. Knouse *et al*., *Science* **361**, 1234–1238 (2018).
20. D. Xu *et al*., *J. Am. Chem. Soc.* **142**, 5785–5792 (2020).
21. K. Kuwabara, Y. Maekawa, M. Minoura, T. Maruyama, T. Murai, *J. Org. Chem.* **85**, 14446–14455 (2020).
22. A. Mondal, N. O. Thiel, R. Dorel, B. L. Feringa, *Nat. Catal.* **5**, 10–19 (2022).
23. D. A. DiRocco *et al*., *Science* **356**, 426–430 (2017).
24. A. L. Featherston *et al*., *Science* **371**, 702–707 (2021).
25. M. Formica *et al*., ChemRxiv [Preprint] (2021). https://doi.org/10.26434/chemrxiv-2021-5714s-v2.
26. C. Bauduin, D. Moulin, B. Kaloun, C. Darcel, S. Jugé, *J. Org. Chem.* **68**, 4293–4301 (2003).
27. T. Kimura, T. Murai, *Chem. Commun.* **2005**, 4077–4079 (2005).
28. T. Kimura, T. Murai, *Chem. Lett.* **33**, 878–879 (2004).
29. A. G. Doyle, E. N. Jacobsen, *Chem. Rev.* **107**, 5713–5743 (2007).
30. D. A. Kutateladze, D. A. Strassfeld, E. N. Jacobsen, *J. Am. Chem. Soc.* **142**, 6951–6956 (2020).
31. A. J. Bendelsmith, S. C. Kim, M. Wasa, S. P. Roche, E. N. Jacobsen, *J. Am. Chem. Soc.* **141**, 11414–11419 (2019).
32. D. D. Ford, D. Lehnherr, C. R. Kennedy, E. N. Jacobsen, *ACS Catal.* **6**, 4616–4620 (2016).
33. K. L. Tan, E. N. Jacobsen, *Angew. Chem. Int. Ed.* **46**, 1315–1317 (2007).
34. H. Xu, S. J. Zuend, M. G. Woll, Y. Tao, E. N. Jacobsen, *Science* **327**, 986–990 (2010).
35. D. A. Strassfeld, E. N. Jacobsen, in *Supramolecular Catalysis: New Directions and Developments*, P. W. N. M. van Leeuwen, M. Raynal, Eds. (Wiley, 2022), chap. 25, pp. 361–385.
36. I. V. L. Wilkinson *et al*., *Angew. Chem. Int. Ed.* **59**, 2420–2428 (2020).
37. A. Babbs *et al*., *J. Med. Chem.* **63**, 7880–7891 (2020).
38. M. Chatzopoulou *et al*., *ACS Med. Chem. Lett.* **11**, 2421–2427 (2020).
39. M. Van Overtveldt, T. S. A. Heugebaert, I. Verstraeten, D. Geelen, C. V. Stevens, *Org. Biomol. Chem.* **13**, 5260–5264 (2015).
40. M. Buti, M. Riveiro-Barciela, R. Esteban, *J. Infect. Dis.* **216** (suppl. 8), S792–S796 (2017).
41. M. Slusarczyk, M. Serpi, F. Pertusati, *Antivir. Chem. Chemother.* **26**, 2040206618775243 (2018).
42. M.-A. Kasper *et al*., *Angew. Chem. Int. Ed.* **58**, 11625–11630 (2019).
43. N. A. Lentini, B. J. Foust, C. C. Hsiao, A. J. Wiemer, D. F. Wiemer, *J. Med. Chem.* **61**, 8658–8669 (2018).
44. Z. Zhao, Q. Zhu, S. Che, Z. Luo, Y. Lian, *Synth. Commun.* **50**, 2338–2346 (2020).
45. J. Xiao *et al*., *Tetrahedron* **74**, 4558–4568 (2018).
46. Y. Xu, Q. Su, W. Dong, Z. Peng, D. An, *Tetrahedron* **73**, 4602–4609 (2017).
47. L. Zhong *et al*., *Asian J. Org. Chem.* **6**, 1072–1079 (2017).
48. Q. Zhu, S. Che, Z. Luo, Z. Zhao, *Synth. Commun.* **50**, 947–957 (2020).

ACKNOWLEDGMENTS

We thank S.-L. Zheng (Harvard University) for determination of the x-ray crystal structure and R. Algera, J. Essman, and H. Sharma for helpful discussions. **Funding:** Funding was provided by National Institutes of Health grant GM043214 (E.N.J.). **Author contributions:** Both authors conceived of the work. K.C.F. designed and conducted the experiments. E.N.J. directed the research. Both authors wrote the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** Crystallographic data for compound **10** are available free of charge from the Cambridge Crystallographic Data Centre under reference CCDC 2155524. All other data are available in the main text or the 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.abp8488
Materials and Methods
Supplementary Text
Figs. S1 and S2
Tables S1 to S11
References (*49*–*56*)

Submitted 1 March 2022; accepted 25 April 2022
10.1126/science.abp8488

By **Christina Petlowany**

Crafty like an engineer

My classmates were certain we needed to use steel. We were designing a wheelchair for a college engineering course and they felt only steel would be strong enough for the handheld levers that would allow the user to propel the chair with a rowing motion. I wasn't so sure. Based on my experience making sculptures with soda cans and creating jewelry with wire, I believed steel would be too heavy and aluminum would be a better option. But the student who most strongly advocated for steel worked at a bike shop; surely I didn't know better, having used metal only for crafts. A few days later, when the hefty, overbuilt steel arm kept flopping down, I felt validated. I had been right—and I wished I had shown more steely resolve in defending my position.

I was a crafty kid. Not crafty like a fox, but crafty to the point that my parents would come home braced for whatever "artistic" explosion I had unleashed that day—origami, painting, clay sculpting, sewing stuffed animals and clothes, and more. But when I enrolled in engineering in college, I put these pursuits aside. Not only was I stretched for time, but I didn't think they were relevant to my academic work—and I hesitated to highlight my "feminine" crafting interests in the male-dominated engineering environment where I already felt like an outsider. I told myself that engineering adequately fed my creative side and I didn't need the hobby.



"Maybe my crafting was something I should embrace rather than hide."

The wheelchair project was a hint that my crafting might be important and relevant, but for the next few years I continued to avoid bringing it up in professional spaces. When I was interviewing for engineering jobs after finishing my master's degree and was asked whether I tinkered in my spare time, for example, I was sure the panelists wouldn't care about my elaborate homemade holiday cards, even though they featured lever action and moving parts. Instead, I muttered about wanting to do more 3D printing. The company extended an offer, so I felt my assumption was confirmed.

My attitude didn't change when I went on to pursue a Ph.D.—until early in the pandemic, when I felt restless and turned to crafting as an outlet. I was making a set of Dungeons & Dragons dice, shimmery blue and purple swirled with gold flakes, as a gift for a friend. While pipetting the liquid resin into the silicone mold, I made an offhand joke to my partner that I was "injection molding"—a standard engineering manufacturing process. I suddenly realized that although resin art is not injection molding in the technical sense, it shares the spirit and probably some skills. Maybe my crafting was something I should embrace rather than hide.

Soon I was seeing more examples of connections between engineering and craft that I had previously overlooked. When working on the wheelchair project, I put my sewing skills to use creating cushioned grips for the handles. The engineering "design kitchen" where my undergrad classmates and I tested our ideas was stocked with inexpensive tools including felt, pipe cleaners, and popsicle sticks—materials that would not be out of place in a craft bin, I now realized. I saw how crafting taught me to persevere when my product didn't match my initial vision and to consider the failed creation a learning and prototyping experience, just as an engineer must.

Since then, I've built crafting back into my free time. I've also stopped hiding it from my colleagues. I mentioned my dicemaking escapades at a robotics conference and broached in a team meeting how we could gain inspiration from an interactive art experience I had recently visited. The responses were consistently positive and constructive—not dismissive or insulting, as I used to fear.

I've grown from a girl who created a makeshift vulpine friend by attaching legs to a stuffed sock and coloring it with red Sharpie to an engineer with valuable skills from my first passion. Perhaps I am crafty like a fox. I am also crafty like an engineer. ■

Christina Petlowany is a Ph.D. student at the University of Texas, Austin. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER