# Science

AaBbCcDdEeFfGgHhIiJjKkLlMm
NnOoPpQq RrSsTtUuVvWwXxYyZz
12      7890

AAAS

# FROM CORE LAB TO YOUR LAB.

*Introducing the NEW easyCyte™ 8HT*

**COMPLETE** - the world's first end-to-end bench top flow cytometry solution.

**OPTIMIZED** - multiparametric kits and assays.

**INTUITIVE** - software and analysis packages.

**POWERFUL** - innovative Guava® technology.

**AFFORDABLE** - fits your bench top and budget.

**Millipore brings the power of flow cytometry to your bench top.**

Imagine running your most complex assays right on your bench top. No core labs. No outside expertise necessary. With Millipore's integrated instruments, assay kits, software and service, you have the end-to-end solution that makes it easy.

**ADVANCING LIFE SCIENCE TOGETHER™**
**Research. Development. Production.**

Make the move to Guava at
www.millipore.com/flowcytometry5

# CONTENTS

**CELEBRATE THE
40TH ANNIVERSARY
OF THE APOLLO 11
MOON LANDING**

The scientific results of that
mission and samples returned
from the Moon, published in our
20 January 1970 issue, are now
freely available with registration
at www.sciencemag.org/apollo11.
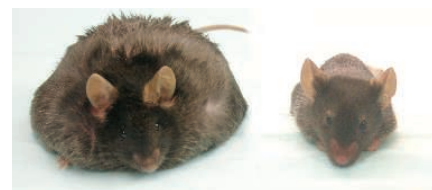
page 256
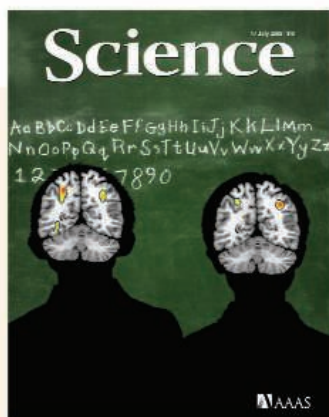
## COVER

Average patterns of brain recruitment, as measured by functional
magnetic resonance imaging, in children with typical reading
development (left) or dyslexia (right) as they sound out printed
words. Left-hemisphere brain regions engaged by typical readers,
the parietal cortex (upper left) and fusiform gyrus (lower left),
are less engaged by dyslexic readers. See page 280.

*Photo illustration: Yael Kats (brain scans, Susan Whitfield-Gabrieli
and John Gabrieli; background, iStockphoto.com)*

page 270

page 300

page 321

# *SCIENCE*ONLINE

**SCIENCENOW**
Avian roots?



**SCIENCESIGNALING**
Growth cone navigation.

## SCIENCEPODCAST
www.sciencemag.org/multimedia/podcast
Free Weekly Show

Download the 17 July *Science* Podcast to hear about the dynamics of rotavirus epidemics, how tiger moths jam bat sonar, NIH's new director, and more.

## ORIGINSBLOG
blogs.sciencemag.org/origins
A History of Beginnings

## SCIENCEINSIDER
blogs.sciencemag.org/scienceinsider
Science Policy News and Analysis

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

## Swimming Through Sand >>

Although composed of solid particles, sand can behave like a fluid. If you had to swim through sand, how would you do it? Would you use your arms and legs for propulsion or would you make your body as compact as possible and try to wiggle and slither your way through? **Maladen** *et al.* (p. 314) used x-ray imaging to study the motion of sandfish lizards as they burrowed into sand. The sandfish lizard does not use its limbs, but instead flattens them against its body and uses large-amplitude traveling wave oscillation of its body to propel itself. Modeling can explain the motion the lizard uses to propel itself through a medium that is neither liquid nor solid.

## A Multiple Photon Pileup

The field of quantum optics began with the observation that two independent photons emitted from a thermal source tend to bunch together. The same is true for any number of bosons, but how do the statistics and correlations evolve experimentally as the number increases? **Aßmann** *et al.* (p. 297) have developed a streak-camera technique that can distinguish the photon number and measure the higher-order correlations between the photons at the detector. The results confirm the predicted "*n* factorial" dependence, showing that the tendency to bunch gets stronger as the number of independent photons is increased.
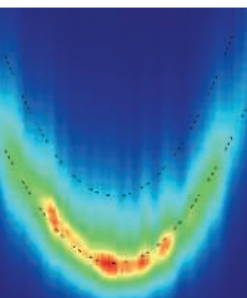
## Living on the Edge

Topological insulators are a recently described state of matter in which the bulk material is an insulator but with a metallic surface state that is protected by the topology of the Fermi surface. **Roth** *et al.* (p. 294; see the Perspective by **Büttiker**) now show that the current flow on the surface takes place in edge states around the boundary of the sample. These are similar to the current transport in high-quality two-dimensional electron gases in high magnetic field, which confirms theoretical work on these materials.

## Shaking Prevents Breaking

Intuition suggests that vibrational excitation of a molecular bond ought to increase the likelihood of its breaking in an ensuing chemical reaction. **W. Zhang** *et al.* (p. 303) observe precisely the opposite outcome in a spectroscopic study of the F + CHD$_3$ reaction. Exciting the C-H stretch leads exclusively to the formation of DF and CHD$_2$ products, in contrast to the more abundant yields of both HF and DF observed in the absence of CH vibration. Though the mechanism underlying this effect remains unclear, the result highlights unanticipated complexity in the reaction dynamics of a relatively simple molecule.

## Dissecting Dyslexia and Learning

Difficulties in learning to read, despite reasonable effort and instruction, form the basis of dyslexia. **Gabrieli** (p. 280; see the cover) now reviews the latest research into the causes of dyslexia. Neuroimaging studies may give early notice of impending dyslexia, and it is hoped that early interventions may lessen the impact of dyslexia. Learning occurs in many settings. Humans uniquely use the formalized settings of schools and curriculum. Infants and children also do plenty of learning outside these settings, often intermingling social interactions. **Meltzoff** *et al.* (p. 284) survey the variety of learning contexts that people experience and discuss how recent advances in neuroscience and robotics are driving a new synthesis of learning.

## Stepping Down

Earth's environment changed markedly over the past 5.2 million years, when a permanent ice sheet has developed in the Northern Hemisphere and the glacial cycle has changed its period from roughly every 40,000 years to the dominantly 100,000-year duration of the past half-million years. One of the biggest questions about these changes is whether they were "threshold" responses to a gradual, uniform cooling trend or whether they represent reactions to discrete episodes of cooling. **Sosdian and Rosenthal** (p. 306) present deep-ocean temperature records from the North Atlantic that show that the cooling happened in distinct steps, at 3 to 2.5 million years ago and at 1.2 to 0.85 million years ago. Combining their record with that of deep ocean water oxygen isotopes allowed the distinction between effects due to global cooling and ice-sheet dynamics.

## Economic Ancient DNA Sequencing

Analysis of ancient DNA is often limited by the availability of ancient material for sequencing. **Briggs** *et al.* (p. 318; see the news story by **Pennisi**) describe a method of ancient DNA sequence retrieval that greatly reduces shotgun sequencing costs while avoiding the many difficulties associated with direct PCR-based approaches. They generated five complete and one near-complete Neandertal mitochondrial DNA genomes, which would have been economically impossible with a simple shotgun approach. Analysis of these genomes shows that Neandertal populations had a much smaller effective population size than modern humans or great apes.

## Moths Battling Bats

Many night-flying insects hear the sonar sounds of attacking bats and take evasive action. Among moths, evasive flight is often accompanied by the production of ultrasonic sounds. Three functions of these sounds have been proposed: to startle the bat, to warn of distastefulness, or to "jam" the

bat's sonar system. **Corcoran et al.** (p. 325) studied a species of tiger moth (*Bertholdia trigona*) that emits a particularly dense series of ultrasonic clicks and the interception behavior of big brown bats (*Eptesicus fuscus*) presented with silenced or sound-producing tethered moths. If the moth sounds evoke startle, naïve bats should initially break off their attacks, but then the bats should habituate to the sounds and complete subsequent attacks. In contrast, if the moth sounds have a warning effect, naïve bats should initially complete their attacks on sound-emitting moths, discover that the moths are distasteful, and refuse to capture them in future trials. Most of the bats in the tests reliably caught the silenced moths but avoided completing attacks on sound-producing moths, with no evidence of increasing or decreasing probability of capture from the first to the last trial, which suggests that the moths effectively jammed the bats' sonar.

## Plethora of Secretory Amyloids

Protein aggregation and the formation of amyloids are associated with several dozen pathological conditions in humans, including Alzheimer's disease, Parkinson's disease, and type II diabetes. In addition, a few functional amyloid systems are known: the prions of fungi, the bacterial protein curli, the protein of chorion of the eggshell of silkworm, and the amyloid protein Pmel-17 involved in mammalian skin pigmentation. Now **Maji et al.** (p. 328, published online 18 June) propose that endocrine hormone peptides and proteins are stored in an amyloid-like state in secretory granules. Thus, the amyloid fold may represent a fundamental, ancient, and evolutionarily conserved protein structural motif that is capable of performing a wide variety of functions contributing to normal cell and tissue physiology.

## The Grim RIPper

Cells can undergo regulated cell death through distinct processes known as apoptosis and necrosis. Regulation of apoptosis is better understood than that of necrosis. In a screen for gene products that participate in control of necrosis in cells treated with TNF (tumor necrosis factor), **D.-W. Zhang et al.** (p. 332; published online 4 June) identified a protein kinase, RIP3. In cells treated with TNF and a caspase inhibitor that inhibits apoptosis, seven metabolic enzymes interacted with RIP3, some of which are associated with mitochondria. Generation of reactive oxygen species was necessary for TNF-induced necrosis, and depletion of RIP3 reduced the generation of reactive oxygen species. Thus, RIP3 may participate in the mechanisms that link energy metabolism with mechanisms of cell death.

## Innate Immunity in the Fly Gut



*Drosophila melanogaster* is an important model system to study innate immunity, being both easy to manipulate and lacking an adaptive immune system. In order to identify genes that regulate innate immunity, **Cronin et al.** (p. 340; published online 11 June) performed an RNA interference screen on flies infected with the oral bacterial pathogen, *Serratia marcescens*. Genes involved in intestinal immunity and regulation of hemocytes, macrophage-like cells critical for phagocytosis and killing of the bacteria, were identified. Several hundred genes conferred either enhanced susceptibility or resistance to bacterial infection. Furthermore, the JAK/STAT signaling pathway was activated in intestinal stem cells after bacterial infection, resulting in enhanced susceptibility to infection, most likely through regulation of intestinal stem cell homeostasis.
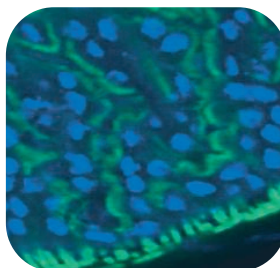
## Resisting Repeats

A set of diseases, including myotonic dystrophy, are caused by the expansion of a simple repeat in genomic DNA, which, when transcribed into RNA, can be toxic to other cellular processes. Ameliorating the effects of this toxic, repeat-laden RNA may also relieve the symptoms of the disease. **Wheeler et al.** (p. 336; see the Perspective by **Cooper**) developed an antisense morpholino oligonucleotide complementary to the expanded repeats found in the myotonic dystrophy protein kinase messenger RNA (mRNA). The morpholino bound the repeats in vitro and displaced the inappropriately bound and sequestered RNA splicing factor, Muscleblind-like 1. In an in vivo mouse model for myotonic dystrophy, local injection of the morpholino corrected a number of cellular defects in muscle, including the alternative mRNA splicing of several genes, among them the muscle-specific chloride channel, CIC1, leading to a marked reduction in the myotonia.

CREDIT: CRONIN ET AL.
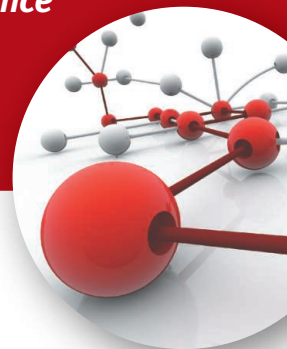
# FROM START TO FINISH, ENDNOTE X3
# KEEPS YOUR RESEARCH ON THE FAST TRACK.

With its world-class speed and track record, EndNote X3 puts your publishing in the lead and keeps you there. Out of the blocks with new features like faster start up times and seasoned favorites like our Cite While You Write™ technology, EndNote X3 is moving fast to deliver the world's most complete bibliographic solution.

It doesn't stop there. EndNote X3 is also meters ahead with a personal EndNote Web account that not only transfers groups swiftly between desktop and Web but also organizes your own publication list for the ResearcherID author community. You'll find new speed in Cite While You Write for Apple® Pages '09, Microsoft® Word, and now OpenOffice.org Writer 3 for Windows. EndNote X3 even sports new options for chemistry styles, multiple bibliographies within a Word document, and more.

Keep the lead, right to the finish line. Put EndNote X3 on your reference team today!

800-722-1227 • 760-438-5526 • rs.info@thomson.com

**Download your free demo or buy online today**
**www.endnote.com**

THOMSON REUTERS

# Earth System Research Priorities

Walter V. Reid is chair of the ICSU Earth System Visioning Task Group, directed the Millennium Ecosystem Assessment, and is Conservation and Science Director at the David and Lucile Packard Foundation, Los Altos, CA.

Catherine Bréchignac is president of ICSU and president of the Centre National de la Recherche Scientifique, Paris, France.

Yuan Tseh Lee is president-elect of ICSU; President Emeritus and Distinguished Research Fellow at the Academy of Sciences located in Taipei; and received the Nobel Prize in Chemistry.

HUMAN-INDUCED CLIMATE CHANGE WAS UNKNOWN OUTSIDE OF LIMITED SCIENTIFIC CIRCLES just 25 years ago, but it has now become the focus of intense national discussions and international negotiations. One chapter in the story of how this issue moved from lab benches to national capitals was recognized by the 2007 Nobel Peace Prize, which was co-awarded to the Intergovernmental Panel on Climate Change (IPCC) for its influential assessments of the state of scientific knowledge. But the story also involves the research itself; in particular, that catalyzed by the Global Environmental Change Research Programmes* and the Earth System Science Partnership. These programs, sponsored by the International Council for Science (ICSU) in partnerships with other international science organizations (www.icsu.org), have helped to catalyze and guide global environmental research for several decades. But it's time to propose new research priorities, and ICSU seeks input through a Web consultation process now under way.

The Global Environmental Change Research Programmes have played major roles in characterizing global biogeochemical cycles, trends and cycles of natural climate change and human-induced warming, and the consequences of those changes for global cycles and human well-being. They have strengthened national research initiatives by forging international research collaborations to study the functioning of the Earth system. But the most pressing research questions are now quite different from those that shaped these programs. The major challenges today include the need to understand complex interrelationships among biological, geochemical, climate, and social systems; the consequences of global change for the Earth system and society and the feedback loops involved; and the science of mitigation, adaptation, and sustainability. Natural sciences should no longer dictate the Earth system research agenda; social sciences will be at least as important in its next phase.

ICSU is spearheading a consultation process in cooperation with the International Social Science Council (ISSC) to renew the focus and framework of Earth system research for the next decade and beyond. The goal is to identify the most urgent research questions and establish the most effective ways to answer them. The process began in 2006 with reviews of the Global Environmental Change Research Programmes. The reviews recommended the creation of a single research framework, an idea also supported by many agencies involved in the International Group of Funding Agencies for Global Change Research. Thus, the focus is now on shaping the new research agenda.

In the past, a small group of scientists would be charged with determining the most pressing research questions. But new communication technologies now allow the wisdom and expertise of a far broader global community of natural and social scientists, technology experts, decision-makers, and citizens to play a role. From 15 July to 15 August 2009, ICSU invites the broad community to shape the Earth system research agenda by contributing ideas and perspectives to a Web forum (http://visioning.icsu.org), as well as by voting on those submitted by others. The results of this online consultation will feed into a September meeting, convened by ICSU and ISSC to distill the input into a set of proposed research priorities.

We know from the experience of the IPCC and the Millennium Ecosystem Assessment that the global research community has the capability to answer many of the most challenging scientific questions about our planet. Now, given the urgent need to confront human-induced global environmental change and the imperative to focus our scientific resources, we need to spread the widest possible net to make sure that the world's scientists will be addressing the questions that are most critical.

**– Walter V. Reid, Catherine Bréchignac, Yuan Tseh Lee**

*Includes the World Climate Research Programme, International Geosphere-Biosphere Programme, International Human Dimensions Programme on Global Environmental Change, and DIVERSITAS.

CREDIT: GETTY IMAGES

## ECOLOGY
### Desperately Seeking Sustenance

Land plants have evolved a variety of specialized adaptations to gather nutrients from unlikely substrates, such as Amazonian trees whose roots grow upward on the bark of neighboring trees. The latest discovery—the snow roots of an alpine plant—comes from 2800 m in the Caucasus Mountains. Onipchenko *et al.* found that the herbaceous plant *Corydalis conorhiza* (a member of the poppy family) produces extensive networks of roots that grow upward and laterally into the snowpack that carpets the high slopes until the July thaw. Isotope experiments showed that these roots, which are anatomically distinct from the normal roots that grow downward into the soil, take up nitrogen directly from the snowpack, thus exploiting a resource that would otherwise disappear down the mountainside during the brief summer. — AMS

*Ecol. Lett.* **12**, 758 (2009).

## BIOPHYSICS
### Assembling Gradient Sensors

Bacteria cluster thousands of transmembrane chemoreceptors at opposite ends of the cell, allowing them to detect and follow food molecule gradients. Might the formation and maintenance of such clusters occur via stochastic assembly? To test this idea, Greenfield *et al.* used photoactivated localization microscopy (PALM) to count single fluorophore-tagged receptors with an optical resolution of 15 nm. They analyzed 1 million receptors and observed that many were present as singletons or small clusters in lateral regions of the cell (shown at left). A mathematical model in which the receptors are inserted randomly into the membrane, but can then be captured and incorporated into existing clusters, accounted for the observed distribution and predicted that the density of new clusters would be highest at a point farthest from a large cluster. Hence, through stochastic assembly, a cell with a large cluster at one pole will form a new large cluster at the opposite pole. Receptor clusters of appropriate size and stability thereby assemble without any specific cellular machinery to position the receptors. — LBR
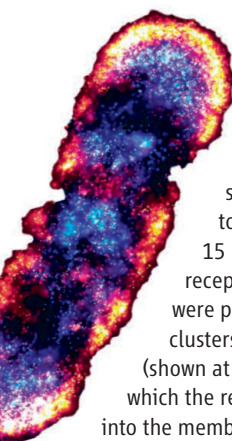
*PLoS Biol.* **7**, e1000137 (2009).

## MATERIALS SCIENCE
### Sponges Tough and Soft

A major challenge in designing biologically compatible implants has been achieving the optimal combination of stiffness, porosity, and toughness (resistance to rupture) for any given local environment. Toward this end, Lee *et al.* used an ionic liquid to condense a dispersion of DNA-coated carbon nanotubes. The ionic liquid efficiently removed bound water from the DNA strands, causing them to form intertwined toroids, which upon drying adopted a porous sponge structure with 50-nm-diameter fibers. The fiber diameter, sponge toughness, and stiffness could be adjusted by soaking the networks in water and then in calcium chloride solution: The calcium ions induced DNA cross-linking. The sponge fibers could then be knotted, braided, and woven into fabric structures. Moreover, the material proved electrically conductive and thus potentially applicable in sensing, energy storage, and mechanical actuation. — MSL

*Angew. Chem. Int. Ed.* **48**, 5116 (2009).

## PSYCHOLOGY
### Correlates of the Gender Gap

In 2003, the Trends in International Mathematics and Science Study (TIMSS) assessed 8th graders on standardized math and science tests. The median score by country for boys was 516 and for girls 506. Nosek *et al.* have studied the relation between gender stereotyping in the general population and student performance on these tests. In their virtual laboratory, any visitor can take an implicit association test (IAT) in any of 17 languages. In more than 500,000 tests collected from 2000 to 2008, roughly 70% of participants tended to associate words representing male with science faster than with liberal arts, and associated words representing female with liberal arts faster than with science. Across 34 countries, the male-female gap measured in the TIMSS correlated with the association of science and male as assessed in the IAT, with one standard deviation in stereotyping equivalent to 6.3 points on the standardized tests. The association of implicit (but not explicit) stereotypes in adults (mean age 27) with national test scores in kids suggests that initiatives aimed at reducing the gap will need to be multifaceted. — BJ

*Proc. Natl. Acad. Sci. U.S.A.* **106**, 10593 (2009).

CREDITS (TOP TO BOTTOM): ONIPCHENKO ET AL., *ECOL. LETT.* **12**, 758 (2009); JAN LIPHARDT

## NEUROSCIENCE

### Detecting Noisy Gradients

Growing neurons are faced with myriad cues as they try to find their designated target. The signals may be soluble or immobile, they may prompt attraction or repulsion, and they may deliver context-dependent messages. Last but not least, any single growth cone interprets its input via a variety of receptors spread across its surface; the growth cone may start or stall, grow quickly or slowly, turn right or left, or reverse course entirely. Mortimer *et al.* have developed a Bayesian model to explain how the growing tips of axons can identify the minute changes in noisy molecular gradients and then interpret them as guidance cues. The optimal strategy for a neuron gives more weight to feedback from receptors that are farther away from the center of the growth cone. Observations of explanted rat neurons facing constructed gradients of signaling ligands in collagen gels showed growth behaviors consistent with this interpretation. — PJH

*Proc. Natl. Acad. Sci. U.S.A.* **106**, 10296 (2009).

## APPLIED PHYSICS

### Liquid Russian Dolls

Oil and vinegar–based salad dressing is a classic example of an emulsion, wherein the droplets of one fluid are trapped inside the bulk of another. Though techniques exist for making higher-order emulsions (with two or more nested droplet layers), they tend either to be inefficient or else to produce droplets that vary widely in size. Abate and Weitz used lithography to fabricate polydimethyl-siloxane devices that can efficiently create uniform distributions of emulsion droplets with up to five nested layers. Single emulsions were prepared using pinned-jet flow focusing—injection of the inner fluid from two side ports into a central stream of outer fluid. To increase the number of emulsion layers, additional injection ports with alternating wettability were added along the length of the flow channel. In order to ensure synchronized droplet formation for triple and higher-order emulsions, the nozzles at each injection port were designed to be slightly narrower than the emulsion arriving from upstream; a new droplet thus formed at the injection port only upon perturbation by an

incoming droplet. The droplets exhibited very narrow size dispersity at all orders, as illustrated by their hexagonal packing when confined in two dimensions. — MSL

*small* **5**, 10.1002/smll.200900569 (2009).

## CLIMATE SCIENCE

### Learning to Share

Governmental representatives from almost every country will meet at the United Nations Climate Change Conference in Copenhagen, in December 2009, in order to attempt to agree on an effective international response to climate change. One of the thorniest and most important questions on the table is how best to determine $CO_2$ emission reduction targets for the various participating countries. This task is rendered more difficult by the asymmetry between developed nations, whose emissions have caused most of the increase in atmospheric $CO_2$ thus far; and less-developed nations, whose emissions have been low in the past but are expected to grow at a faster than average rate in the future. Chakravarty *et al.* propose that national reduction targets, for the near term, be based not on per capita emissions, but on the net excess emissions from the individual high emitters that are found in every country. This approach has the advantages of treating equally all those with the same emissions, regardless of nationality, and of not specifying how any nation meet its responsibilities for reducing $CO_2$ emissions. — HJS

*Proc. Natl. Acad. Sci. U.S.A.* **106**, 10.1073/pnas.0905232106 (2009).

## MICROBIOLOGY

### Fingers or Toes?

Countless hours have been spent on scrutinizing the morphological subtleties of planktonic organisms, particularly in trying to match shapes to species and to reconcile both with the huge genetic diversity; sometimes, the disconnect can be profoundly misleading. By tracking individuals in culture-well plates, Pizay *et al.* noticed that dinoflagellates changed shape in striking ways. *Ceratium ranipes* grew rigid chloroplast-filled fingers by day and became relatively lethargic, whereas at night, they absorbed the appendages and became more active. Why? One possibility is that the daytime appendages allow the organisms to maximize photosynthesis at the surface, and nighttime absorption allows them to sink a little, swim a little faster, and escape predator pressure. — CA

*Protist* **160**, 10.1016/j.protis.2009.04.003 (2009).

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

## Norway's Summer Skies

Astrophysicist Donald Olson and colleagues at Texas State University, San Marcos, have been finding lots to study in the paintings of Norwegian artist Edvard Munch. Their latest accomplishment: identifying the celestial objects in three canvases Munch painted in Åsgårdstrand, Norway. After locating a memoir that placed Munch in Åsgårdstrand in August 1893, the researchers traveled to Norway to find the exact sites of the paintings.

In *Starry Night* (right), scholars in the past have identified the bright star as Venus. But Venus was out of sight then, the Texans say. A photo (left) they took from the same perspective shows it had to be Jupiter. They also figured out, with the help of 19th century photos of the town, that the vertical white line in the trees, which some have identified as a hidden moon and its reflection, was in fact a flag pole with a ball on top.
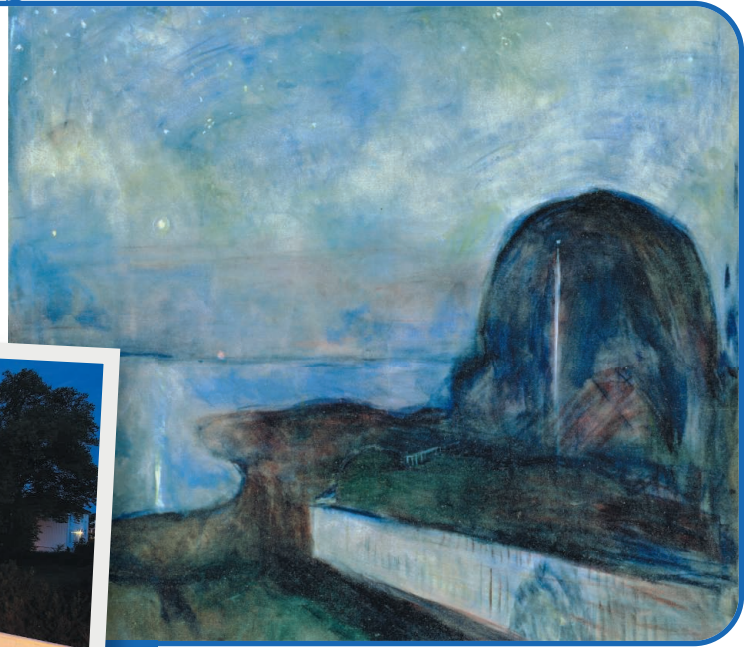
## RUPTURE AND REUNION

Marie Csete, chief scientific officer at the California Institute for Regenerative Medicine (CIRM), will depart at the end of July after only 15 months on the job. Csete won't disclose specific reasons for leaving, but it has been reported that she wasn't able to get CIRM Director Alan Trounson and board chair Robert Klein to listen to her.

**Q: Before coming to CIRM, you divorced your husband, John Doyle, a math professor at the California Institute of Technology (Caltech). Is it true you got divorced in order to take this job?**
**M.C.:** It's really true. I really turned my life upside down to do this. I think when they passed all these fair political practices laws in California, they were never meant for scientific conflicts of interest; they were really meant for politicians. I fully expected that if I took the job and Caltech came up, I would just walk out of the room [to avoid a possible conflict of interest should Caltech apply for CIRM money]. But that wasn't good enough for the lawyers. [The divorce] didn't change our relationship, which was still a commute. [Csete moved from Emory University in Atlanta, Georgia, to San Francisco; Doyle lives in Los Angeles.] It actually got worse because by law, he wasn't allowed to take a salary from Caltech for a year. So he had to retire for a year with no salary.
**Q: How did your husband feel about this?**
**M.C.:** He thought it was worth the sacrifice; he thought I was the right person for the job.
**Q: Can you say anything about why you're leaving?**
**M.C.:** There are principles very important to me that I knew after trying that I could not change from within. The only way was to step away and talk to people on the way out.
**Q: What were your major accomplishments?**
**M.C.:** I've established very strong relationships between the science office and grantees, … allowing scientists to be doing some real science rather than just writing RFAs [requests for applications] all the time.
**Q: A lot of people think it's a shame you're leaving.**
**M.C.:** I've gotten a lot of e-mails—more than 100—from people. Some of them I'm afraid to open because it made me cry. … Very sad e-mails.
**Q: Future plans?**
**M.C.:** I don't know. I want to look for a place where I can make a real difference. I would prefer to be close to my husband. We'll get remarried soon.

## Losing Battle in Britain

With the highest teen birth rate in Europe, the United Kingdom is anxiously seeking ways to stem the tide. But researchers say a program to reduce pregnancies among teens seems to have had the opposite effect.

Scientists led by Meg Wiggins at the University of London studied how the Young People's Development Programme, which ran at 27 locations in England from 2004 to 2007, affected more than 2300 "at risk" 13- to 15-year-old boys and girls. They compared data on sexual behavior, drug and alcohol use, and school suspensions with statistics for students from 27 comparable areas.

The after-school program offered 6 to 10 hours a week of tutoring, sex education, health services, art classes, and career counseling over a year. The results, reported last week in the *British Medical Journal*, showed "significantly" more pregnancies in the intervention group than in the comparison group: 16% versus 6%. Girls in the program also had sex earlier and were more likely to expect to be mothers by the age of 20. The program had no discernible effect on boys.

The researchers suggest that girls might have been influenced by exposure to risky peers or even just by being labeled "at risk." Curbing teen fertility is an uphill struggle, says evolutionary psychologist David Buss of the University of Texas, Austin: "Teen women today are simply doing what their maternal ancestors did over human evolutionary history."

NOMINATIONS

# White House Taps Former Genome Chief Francis Collins as NIH Director

President Barack Obama's announcement last week that he had chosen Francis Collins to lead the National Institutes of Health (NIH) did not come as a big surprise. But it ended months of speculation and ignited a volley of flattering remarks from researchers and biomedical groups. "Francis is one of the most accomplished scientists and scientific leaders of his generation. … Having worked with him for many years, I am sure that he will rise to the unique challenges of this job," said Elias Zerhouni, who resigned as NIH director last fall.

Collins is known as a skilled administrator and excellent communicator. Over 15 years, he built a new center at NIH into one of the most visible and innovative institutes. When he stepped down as leader of the National Human Genome Research Institute (NHGRI) last year, he was already considered a leading candidate to run NIH, the $30 billion parent agency.

Although few would disagree with a White House press notice saying that Collins's work "has changed the very ways we consider our health and examine disease," Collins does have critics. Some question his support of "big biology" in the genome project portfo-lio—with timetables and planned targets—and some are concerned about his outspoken Christian faith. He raised eyebrows, for example, when he recently launched a Web site, BioLogos, expanding on his 2006 book explaining how he reconciles his faith with the science of evolution (see sidebar, below).

Biomedical scientists are pleased, however, to have a permanent leader at NIH, which has been run by an acting director, Raynard Kington, since October. The agency is staggering under an unprecedented number of grant applications seeking to share in NIH's $10 billion windfall from the economic stimulus package. When that money runs out in 2011, it's unclear what will happen to stimulus-funded scientists. If it comes to a crunch, they could benefit from having an internationally renowned genome scientist as a spokesperson.

Collins, 59, grew up home-schooled during his early years on a farm in rural Virginia and later earned a Ph.D. in physical chemistry and

**Back to Bethesda.** Francis Collins appears to be a shoo-in for NIH director.

## QUESTIONS ABOUT THE LANGUAGE OF GOD

Although many scientists say geneticist Francis Collins will make a superb director of the National Institutes of Health (NIH), not everyone is celebrating. A discussion about whether Collins's very public religious views will influence his leadership of NIH played out on blogs early this spring and again in the past week. There seems to be little evidence for such worries, but they persist.

Collins has written that his beliefs played a role in the 2000 White House press conference to announce the draft sequence of the human genome, when President Bill Clinton called the human DNA sequence "the language in which God created life." In 2006, Collins wrote a book, *The Language of God: A Scientist Presents Evidence for Belief*, that describes his religious conversion at the age of 27 and how he reconciles this with the science of evolution. Richard Dawkins, the biologist and prominent anti-religionist, feuded with Collins for mixing science and faith.

This spring, Collins raised hackles again when he and several other scientists launched a foundation and Web site, BioLogos, which claims that it "emphasizes the compatibility of Christian faith with scientific discoveries about the origins of the universe and life." Funded by the Templeton Foundation, which supports projects at the intersection of science and religion (including at AAAS, *Science*'s publisher), BioLogos answers faith-related questions and links to a blog by its founders.

As weeks passed with Collins the rumored nominee to head NIH but no announcement, some speculated that BioLogos might be an obstacle. One prominent critic, Paul Z. Myers, a biologist at the University of Minnesota, Morris, who runs the anticreationist blog Pharyngula, faults Collins for suggesting that altruism cannot be explained by evolution and instead came from God. "Collins has got some big gaps in his understanding of the field of evolutionary biology," Myers says. In comments this spring on Pharyngula, others fretted that Collins's beliefs could influence his decisions on topics such as stem cells and sex research.

But others have pointed out that Collins's record as director of the genome institute doesn't support such fears. And some scientists active in the anticreationist movement approve of his attempts to reach out to the faithful. Evolutionary geneticist Wyatt Anderson of the University of Georgia in Athens says he read Collins's book, and "I get the picture of a very rational scientist." Josh Rosenau, public information project director of the National Center for Science Education in Oakland, California, says: "It's very useful to have scientists out there like Francis Collins to talk about their beliefs and why they don't see them as in conflict with science."

As of last week, Collins is now only "minimally involved" with BioLogos, says his wife, Diane Baker, a BioLogos board member. She says he plans to step down from the foundation once the Senate has confirmed his nomination and that he will decline any speaking engagements or efforts to promote BioLogos. **–J.K.**

an M.D. At the University of Michigan, his team made early gene-hunting discoveries, finding genes for Huntington's disease and cystic fibrosis. In 1993, he joined the human genome center, taking over from DNA structure co-discoverer James Watson.

Collins steered the ramp-up of the Human Genome Project, crafting pioneering agreements that required scientists to share genomic data freely. He led a sequencing race against a private effort headed by J. Craig Venter that culminated in a tie to finish a rough draft of the human genome in 2000. Since then, Collins has steered follow-on efforts, including the HapMap, which has fueled the search for genetic risk markers for common diseases.

Throughout his career, Collins has charmed Congress and the public. He helped push a law through Congress last year that bars discrimination based on genetics. When he resigned from government in 2008, he said he wanted to write a book about personalized medicine but soon thereafter penned an op-ed piece endorsing Obama. He later joined the president's transition team and this year tried to help religious groups come to terms with Obama's order easing limits on the use of federal funds to study human embryonic stem cells.

Even Collins's biggest fans say that their star will need to carve out a larger role. Collins did "a fabulous job as NHGRI director," says geneticist Aravinda Chakravarti of Johns Hopkins University in Baltimore, Maryland. But now, Chakravarti says, "he will need to understand, feel, and anticipate the interests of a much broader constituency," including small-lab investigators in fields such as infectious diseases and cell biology who have felt threatened by the big projects that Collins has championed.

Collins will also face concerns that the payoff from the Human Genome Project has been oversold. So far, the search for risk markers for common diseases arguably has found little that could be applied to patients. "I do think one of Francis's tasks will be to set high hopes for genomics but also to manage expectations," especially to convey realistic time frames, says Robert Cook-Deegan, a medical ethicist at Duke University in Durham, North Carolina, and former NHGRI staffer. Collins has had his own DNA "SNP-chipped" for genomic markers but has said

## Online
### sciencemag.org
Podcast interview with author Jocelyn Kaiser.

that so far these markers seem most useful for finding new drug targets, not predicting risks. In his new book, due out in 2010, Collins is expected to maintain that the payoff of molecular genetics is coming.

Collins's plate is loaded with controversial issues. Among others, the next NIH chief must craft new conflict-of-interest regulations for grantees and look at a possible restructuring of the entire NIH operation. The "obvious acute issue," says molecular biologist Keith

Yamamoto of the University of California, San Francisco, "is the stimulus money and concern about the 2011 cliff." Yamamoto has been involved for many years in reforming peer review and was interviewed for the NIH director position.

Obama has sent Collins's nomination to the Senate for review by the Health, Education, Labor, and Pensions Committee. Collins isn't commenting to the press, but supporters say they hope he can be confirmed before the Senate breaks for recess on 7 August. **–JOCELYN KAISER**

## OBAMA NOMINEE
# Geophysicist McNutt Named to Lead U.S. Geological Survey

Even before she was officially nominated last week to be the next director of the U.S. Geological Survey (USGS), Marcia McNutt was already angling for broader responsibilities. After her prospective boss, Interior Secretary Ken Salazar, told her that he hoped to elevate science throughout the department, McNutt replied: "Then why don't you make the director of the Geological Survey your science adviser?"

When Salazar said yes, he put the 57-year-old geophysicist in line to make history twice—as the first woman to lead the 130-year-old survey and as the department's first science adviser. "Now science advice is going to have a more direct way to be communicated to the other agencies," she says, adding quickly that nothing will happen unless the U.S. Senate confirms her appointment.

Such directness and initiative could be hallmarks of her tenure at the 8800-employee, $1.1-billion-a-year agency, an amalgam of geologists, biologists, and hydrologists. As science adviser, she'll also be working closely with other scientific branches of the department, such as the National Park Service and the Fish and Wildlife Service.

McNutt has made a habit of breaking new ground. She was the first female physics

**New duties.** USGS nominee Marcia McNutt would also act as science adviser to the Interior's Ken Salazar.

major at Colorado College, the first female lifeguard for the city of Minneapolis, and, she guesses, the first woman to train in underwater demolition with Navy SEALs. Her career in tectonophysics has taken her from the ocean island volcanism of French Polynesia to the uplift of the Tibetan Plateau. She began her professional life with 3 years at USGS before joining the Massachusetts Institute of Technology in 1982. Since 1997, she has been president and CEO of the Monterey Bay Aquarium Research Institute, a $40-million-a-year, 200-employee oceanographic research institution in Moss Landing, California, that focuses on the nearshore of Monterey Bay.

"She brings an incredible skill set to the job," says geophysicist Mary Lou Zoback of Risk Management Solutions in Newark, California, who was formerly with USGS. "Marcia's nomination clearly affirms that the Geological Survey is a science agency that should be led by a scientist." Zoback predicts that McNutt will work closely with two other recent Administration appointees, marine ecologist Jane Lubchenco of the National Oceanic and Atmospheric Administration and Secretary of Energy Steven Chu.

**–RICHARD A. KERR**

CREDIT: TODD WALSH © 2003 MBARI

**Across the continent.** Entire mitochondrial genomes now exist for these four Neandertal bones and two more, suggesting low genetic diversity in this extinct human.

El Sidrón Cave

Neander Valley

Vindija Cave

Mezmaiskaya Cave

ANCIENT DNA

# Sequencing Neandertal Mitochondrial Genomes by the Half-Dozen

Fourteen years ago, sequencing just a few hundred bases of mitochondrial DNA from a Neandertal drew applause worldwide. Ancient DNA studies have come a long way since then. On page 318, a team led by Svante Pääbo of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, describes a new technique the team used to decipher the entire mitochondrial genomes from five of these extinct humans. These genomes show relatively little genetic diversity among Neandertals scattered across Europe and Russia.

"This is an important step for ancient DNA," says Eddy Rubin, whose lab at the Lawrence Berkeley National Laboratory in California did some of the first work sequencing Neandertal nuclear DNA. Adds Eske Willerslev, who studies ancient DNA at the University of Copenhagen in Denmark, the new technique "provides a solution to a technical problem: … how to target specific regions of interest."

Ancient DNA is difficult to sequence because genetic material degrades over time into small fragments just tens of bases long and errors are often introduced into the aging sequence. Moreover, 99% of fossil DNA tends to be contaminating sequence from microbes and fungi that have infiltrated the decaying bone. Three years ago, researchers began sequencing all of the DNA in Neandertal samples, then separating out what looked like human sequence from the mix. New sequencing technologies made the project affordable enough to go after the entire Neandertal nuclear genome (*Science*, 17 November 2006,

p. 1068), which was announced in February (*Science*, 13 February, p. 866).

Some of that project's DNA came from mitochondria, and researchers assembled it into the first entire Neandertal mitochondrial genome in 2008. Sequencing another mitochondrial genome that way would cost as much as $400 million, says Pääbo graduate student Adrian Briggs. That would make comparing multiple Neandertal genomes—the only way to understand the population size and structure of our closest relative—prohibitively expensive. So Briggs came up with a better way. His new cost: about $8000 per mitochondrial genome.

The approach uses probes that recognize and isolate only Neandertal mitochondrial DNA from all the contaminating DNA in a sample. Thus, Briggs winds up sequencing just the material he's looking for.

To do this, he first attaches short sequence tags to all the pieces of DNA in his sample, creating a DNA "library" that can be copied to ensure there will be enough ancient DNA for future use. Using the already-sequenced Neandertal mitochondrial genome as a template, Briggs designed 574 probes to cover the entire mitochondrial genome. When a probe links up with its target sequence in a mass of DNA fragments, an enzyme copies the rest of the DNA in that piece many times over. Thus, Briggs was able to isolate the whole mitochondrial genome and sequence it using the latest high-throughput sequencing technologies. Others have used similar approaches, but Briggs tailored his for the short, degraded fragments found in fossils.

Briggs and his colleagues sequenced 16,565 mitochondrial bases extracted from bones from Spain, Germany, Croatia, and Russia (see map) and analyzed those genomes along with the one sequenced earlier, which comes from a long bone fragment from Croatia. The bones ranged in age from 38,000 to 70,000 years old. The team also compared the ancient DNA with mitochondrial genome data from about 100 modern individuals.

Briggs and postdoc Jeffrey Good found 55 places out of the 16,565 bases where the mitochondrial genomes varied across the six ancient samples. On average, they found 20 differences between any two samples. In modern humans, about 60 differences exist between any two samples, making Neandertals about one-third as diverse. That isn't unexpected given that humans come from across the globe and that the Neandertals were confined to Europe and Russia, notes John Relethford, a biological anthropologist at the State University of New York College at Oneonta. The results call into question earlier suggestions that Neandertals were divided into separate, regional populations, but more data are needed to be sure, says Briggs.

By the Max Planck group's calculations, this diversity translates into the equivalent of at most 3500 breeding Neandertal females, or up to 7000 including males, lower than previous rough estimates of about 10,000. This so-called effective population size is far less than the actual population and represents a theoretical attempt, based on the population's genetic diversity, to quantify the number of individuals who are breeding at any given time. For example, Anders Götherström of Uppsala University in Sweden calculates that although the Swedish population numbers 9 million, the effective population size is about 100,000; he estimates that the 3500 might translate very roughly into about 70,000 Neandertals.

"Low population size may have been a general aspect of Neandertal biology," says Briggs. With relatively few individuals, the species may have been more vulnerable to extinction from climate change or competition from our ancestors, he adds.

Such a small number "is not too surprising," says ancient DNA expert Alan Cooper of the University of Adelaide in Australia, as archaeological evidence had been pointing toward this. Cooper and others caution that analyzing mitochondrial DNA has limits as it is "in effect 'one gene' because all its genes are so tightly linked." Thus, says Cooper, "this is just one view of Neandertal evolutionary history."

–ELIZABETH PENNISI

INDIA

# Lunar Survey Spacecraft Develops an Attitude Problem

**NEW DELHI**—India's first moon probe, Chandrayaan-1, has suffered a critical malfunction that jeopardizes the remainder of the mission. The spacecraft, which entered lunar orbit last November, can no longer orient itself with high precision. "Its pointing accuracy has been compromised," says a mission engineer who asked for anonymity.

Chandrayaan-1 achieved all of its mission objectives before the malfunction was detected in May, says G. Madhavan Nair, chair of the Indian Space Research Organization (ISRO) in Bangalore. It was a "dream run" until then, Nair told *Science*. Some foreign scientists with instruments aboard Chandrayaan-1 concur that the probe performed well. "The data … are unique and reveal a new moon," says Stas Barabash of the Swedish Institute of Space Physics in Kiruna, whose Sub keV Atom Reflecting Analyser is investigating the interaction of the solar wind with the lunar surface. The spacecraft is con-



**Stunning images.** Chandrayaan-1 has returned excellent data despite a glitch, says G. Madhavan Nair.

tinuing a search for water ice, gathering data for a three-dimensional lunar atlas, and mapping minerals.

The first inkling of something amiss came on 19 May, when the satellite was raised from a lunar orbit 100 kilometers above the surface to a 200-kilometer orbit—allowing for greater stability and easier maneuvering. ISRO did not disclose the pointing failure; instead, it announced that "after successful completion of all major mission objectives," the higher altitude would enable further studies on gravity anomalies and imaging a wider swath of the lunar surface. Nair denies that ISRO sought to hide the problem with the $100 million robotic mission, saying there was no need to go public since there was "no degradation or deterioration in the mission."

The broken instrument is a star sensor, which orients the probe. Its failure is "disheartening," says George Joseph, director of the Center for Space Science and Technology Education in Asia and the Pacific in Dehradun. "The precision with which the spacecraft was maneuvered into the moon orbit was in itself a fantastic achievement," says Joseph, who helped design the mission.

Mission scientists say Chandrayaan-1 hit most of its scientific milestones. "Visually arresting" images from NASA's Mini-SAR radar of craters in permanent shadow "will be extremely useful in unraveling the complex geological history of the moon," says principal investigator Paul D. Spudis of the Lunar and Planetary Institute in Houston, Texas. "Chandrayaan's achievements are quite remarkable," adds Carlé Pieters of Brown University, principal investigator of the Moon Mineralogy Mapper. The higher orbit will require the team to "replan" operations, Pieters says, "but overall the tradeoffs will probably result in an equally strong science product."

The payload that will suffer most is India's Lunar Laser Ranging Instrument, designed to measure altitude variations within a 5-meter accuracy. It was designed to operate at a 100-km orbit; at 200 km, the return signal may be too weak, says a mission specialist. Also unclear is whether projects tied to x-ray and near-infrared spectrometers will be completed, says Christian Erd of the European Space Agency in Noordwijk, Netherlands.

"There is nothing to be worried or alarmed about," insists Nair. "The mission is almost over." Indeed, the global scientific team will meet in late August or early September in India to decide whether to keep Chandrayaan-1 going or guide it to a controlled crash before it completes its nominal life of 2 years.

**–PALLAVA BAGLA**

CREDITS (TOP TO BOTTOM): D. GRIMM/SCIENCE; PALLAVA BAGLA

---

## *ScienceNOW*.org

### From *Science*'s Online Daily News Site

**Cat Purrs Evoke Baby Cries**
There may be more to a cat's purr than meets the ear. A new study reports that our feline friends modify their signature sound when seeking food, adding a higher-frequency element that exploits our sensitivity to infant wails—and thus making it harder to ignore. http://bit.ly/pVDMd
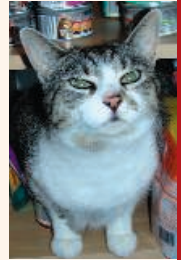
**Don't Blame Birds for 1918 Flu**
It has become almost common wisdom that the virus that caused the 1918 flu pandemic was an avian strain introduced into the human population shortly before the pandemic erupted. But a new study disputes that hypothesis, arguing instead that genes of the 1918 virus had circulated in mammalian hosts, most likely pigs and humans, for several years before 1918. Multiple gene-swapping events brought them together in a single killer strain, say the researchers; improving surveillance in humans and in swine could, in the future, give scientists an early alert to such events. http://bit.ly/3mg2l9

**Swearing Eases the Pain**
You just stubbed your toe, hard, on the corner of that stupid table again. What's the first thing out of your mouth? If it's something you wouldn't see printed in a family newspaper, you may actually be doing yourself a favor. Foul language may be decreasing your pain, according to a new study. http://bit.ly/Cg7Qk

**Flexible Fibers Act Like Cameras**
Picture a wall that stares back at you. Or a uniform that shows a soldier a 360-degree view of a battlefield. Both scenarios are possible courtesy of a new generation of flexible, translucent fibers developed by researchers at the Massachusetts Institute of Technology in Cambridge. These so-called multimaterial fibers can turn incoming light waves into images without the need for a camera lens. And unlike fiber-optic cables, they can transmit images that have been captured across their entire length. http://bit.ly/GB6oi

Read the full postings, comments, and more on **sciencenow.sciencemag.org**.

## ARCHAEOLOGY

# Roundup of Utah Collectors Stirs a Debate on Enforcement

**BLANDING, UTAH**—Last month, after 16 residents from this town in southern Utah were arrested and accused of stealing prehistoric Indian artifacts from public and tribal lands, nobody was more surprised than Teri Paul, director of the state's Edge of the Cedars State Park Museum, which features archaeological exhibits on the area's Native American cultures. "I guess I just got it into my head that locals weren't doing this anymore."

The town was the center of a similar raid back in the mid-1980s. Old traditions die hard, though. Blanding, like numerous other communities in the Four Corners region, has a long-standing tradition of digging up Native American ruins for recreation and profit. The habit took hold during American archaeology's embryonic period between the 1800s and 1920s—ironically, a time when the Smithsonian Institution and others sponsored expeditions to the Southwest and paid locals to find artifacts. Federal laws now make it a crime to take such trophies from public lands and Indian reservations, but the practice of "pothunting" continues, fueled by a lucrative black market in antiquities.

This latest crackdown is the culmination of a 2.5-year-old undercover investigation by the Federal Bureau of Investigation (FBI) and the Department of the Interior that netted 24 alleged looters in the archaeology-rich region. It is renowned for Chaco Canyon's Great Houses and Mesa Verde's cliff dwellings, both grand vestiges of the Anasazi, a farming civilization that flourished from 500 C.E. to 1300 C.E. The dragnet on 10 June, *Science* has learned, extended to art dealers and collectors in a dozen cities from Tucson, Arizona, to Santa Fe, New Mexico. Agents searched homes and businesses and seized personal files and computers. Two members of a family in Blanding accused of looting last week pleaded to charges and surrendered a collection of artifacts.

The case has aroused strong passions among Southwestern archaeologists, prompting many to take sides on how best to cure pothunters of their destructive urge. Some strongly support the federal action; others take a jaundiced view of the heavy-handed police tactics and argue that such an approach will not deter determined looters, especially those who come from communities where a subculture of pothunting stubbornly persists.

Government archaeologists involved in the investigation paint an emerging picture that is as lurid as it is far-reaching. "We're talking widespread, systemic destruction of archaeological sites," says Emily Palus, a Bureau of Land Management (BLM) archaeologist based in Washington, D.C. Many of the items were dug out from Indian burials, such as a turkey-feather blanket, a copper bracelet, and a pair of ancient sandals associated with the Anasazi.

Scientists say looting of these sites is akin to tearing pages out of a history book. "If you rip out enough pages, pretty soon the book doesn't make sense," says Jerry Spangler, executive director of the Colorado Plateau Archaeological Alliance, a Utah-based antiquity preservation group.

Enforcement of antitheft laws has been lax in the past, according to many Southwestern archaeologists. But Spangler believes the recent sweeps finally demonstrate "that federal authorities do consider pothunting as a serious crime." The message of the raids is unmistakable, particularly in the way they were carried out, with FBI SWAT teams descending on homes with guns drawn and placing arrested suspects in handcuffs and leg irons, says Richard Wilshusen, an archaeologist who teaches at Colorado College in Colorado Springs: "It's the OK Corral. It's the Clanton Gang finally being taken on."

Among the Blanding citizens swept up in the 10 June raid was James Redd, the town's prominent physician, who killed himself the next day. (Another suspect from Durango also took his own life a week later.) Redd's death has since triggered an angry backlash against the federal raids, with both of Utah's Republican senators calling the investigation "overkill" and demanding that Congress initiate a probe into FBI's antipothunting operation. The backlash may feed a growing debate among experts over the use of punishment—rather than education—as a means of changing public attitudes.



**After the bust.** Interior Secretary Ken Salazar (*left*) announces the arrest of Utah residents for taking artifacts from federal and Native American lands.

### Canyons of riches

The biggest challenge, many archaeologists say, is conveying the importance of archaeological protection for scientific purposes. That's difficult enough when Indiana Jones and a treasure-hunting aura popularly define the field. The challenge is made even harder in a place like Four Corners, which is believed to contain the highest density of archaeological sites in the country, if not the world. For example, after news spread of the raids in Blanding, one resident (whose in-laws were among those arrested) told *The Salt Lake Tribune*: "It's just something everyone does in Blanding. There are artifacts everywhere."

Overall, a history dating back at least 10,000 years can still be widely found across the landscape, from rock-art panels and arrowheads to collapsed pueblos (above-

No context. The polychrome bowl (*right*) is classed as having an unknown origin, as many recovered artifacts will be.

ground dwellings) and rock shelters. In San Juan County alone, where Blanding is located, more than 25,000 archaeological sites have been inventoried. The Edge of the Cedars museum was built directly adjacent to the stone walls and kivas of a 1000-year-old Anasazi village. Paul, noting that 92% of the county is federal land, much of it rugged wilderness, says, "In any given canyon, you'll find site after site."

But even many of these sites are being "vacuumed" of surface artifacts, such as potsherds and arrowheads, says Paul. That eliminates vital clues to possible dwellings and other larger sites nearby, which are often buried underground. Experienced pothunters, using shovels or, worse, a backhoe, take the damage to another, incalculable level. "Digging and removing artifacts destroys their context, and it limits the information that can be obtained from a site about previous cultures," says Laird Naylor, a BLM archaeologist in the Monticello Field office, whose jurisdiction covers San Juan County, Utah.

Spangler points out that preservation is also important because methods and science are always improving. "You want to be able to save these sites for the archaeologists that come after us 20 and 50 years from now, because they're going to be so much more advanced than us."

Archaeologists admit that the reasons for site protection are poorly conveyed to the general public. That's why Jonathan Till, a contract archaeologist with Abajo Archaeology in Bluff, Utah, argues that getting pothunters to change their ways will happen through education, not prosecution. He says BLM is failing miserably because it lacks resources and labor power. Till notes that BLM oversees 800,000 hectares of public land in San Juan County and has only two archaeologists and one ranger. "That's just absurd," he says. "Those people are stretched way too thin to be able to do their jobs effectively and to be able to act as educators. They just cannot do that."

Brian Quigley, acting manager for BLM's Monticello field office, agrees that it would be nice to have more boots on the ground: "We do the best we can with the resources we have," he says. The situation is similar in BLM offices across the state of Utah. For example, just north, in the Price field office, an area also chock-full of archaeological sites, including the world-famous Nine-Mile

Canyon under BLM's watch, Blaine Miller is the sole archaeologist. "I've been the only one here for the past 25 years," he says.

Given this shortage of labor power, archaeologists are divided over the best way to stamp out pothunting. Winston Hurst, a respected independent archaeologist who consults on academic and industry projects and lives in Blanding, criticizes the federal

## ARTIFACT LOCATIONS

1. Mancos Mesa
2. Bullet Canyon
3. Alkali Ridge
4. Baby Mummy Cave

Looted sites. Culminating a 2.5-year investigation, a coordinated sweep seized objects at many locations around Blanding, Utah.

raids as "bizarre theater" in which "the symbolism of the way it was carried out ends up trumping the substance of why they're doing it in the first place." Hurst says the only way to conquer archaeological looting is through "intelligent discourse, and by treating people with respect." But if "you're going to punch people in the nose, there's absolutely no open-mindedness, nothing but a fight, and once it's a fight, they're going to retrench in opposition to you."

Other Southwestern archaeologists are also inclined to take a softer approach. "No archaeologist likes looting, but it seems like so much overkill to do this to normal people," says Catherine Cameron, a professor of archaeology at the University of Colorado, Boulder. "They surrounded houses early in the morning, like it was a drug bust. These are not bad people. You don't need to do that to people like they are dangerous criminals."

But Kevin Jones, Utah's state archaeologist, has little sympathy for the Blanding residents who were implicated in robbing graves. "Maybe upstanding citizens are not used to being arrested that way; maybe now they will think twice about committing that kind of crime." **–KEITH KLOOR**

Keith Kloor is a writer in New York City.

## From the *Science* Policy Blog

Congressional spending panels last week looked favorably on basic research programs at **the Department of Energy (DOE)** but were skeptical about innovations proposed for 2010 by the Obama Administration. House of Representatives and Senate appropriators came very close to the president's $4.94 billion request for the Office of Science. But they showed **little enthusiasm for Energy Secretary Steven Chu's plan to fund eight large energy research centers**, with the House approving $35 million of the $280 million request and the Senate suggesting a way to spend $44 million.

At a meeting at the U.S. National Academy of Sciences last week, **NASA Associate Administrator Edward Weiler** warned that the shrinkage of NASA's planetary budget from $3 billion to $1.5 billion in the past 4 years means that **"we no longer have a viable Mars program."** Weiler announced an unprecedented agreement with the European Space Agency to conduct a joint program of Mars missions.

The **University of California** released a plan last week that will shave $184 million from the university's projected $813 million shortfall in state funding over the next 2 years. It calls for **furloughs to be scaled according to pay grade**—from 11 days, equivalent to a 4% cut, for those making less than $40,000 a year to 26 days, or a 10% cut for those making more than $240,000. Employees funded entirely from nonstate sources would be exempt.

**Maria Leptin**, the new director of the European Molecular Biology Organization, wants to review how to balance EMBO's reliance on its journals for revenue with the scientific community's desire to make the journals open-access.

**Health care workers should be first in line for inoculation** when vaccines against the **swine flu** virus are ready and approved, an expert panel at the World Health Organization concluded in a meeting last week.

For **a full Q&A with Leptin** and other science policy news, go to **blogs. sciencemag.org/scienceinsider**.
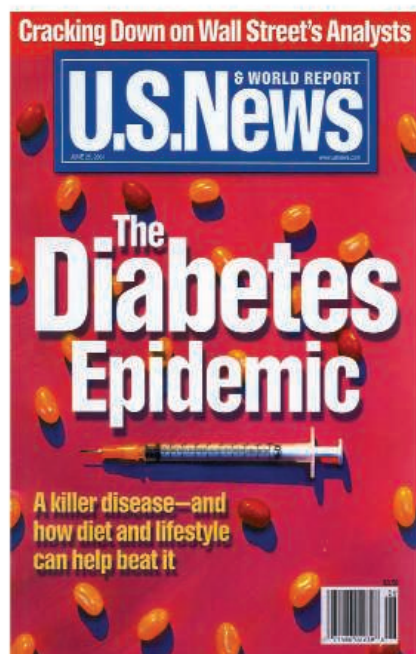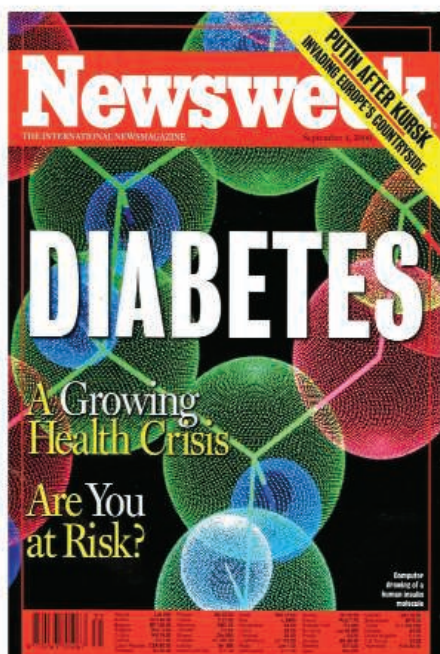
# Prosperity's Plague

**Researchers have linked a growing number of chronic diseases to the metabolic disorder known as insulin resistance; two general theories have emerged about its mechanism**

Welcome to the age of insulin resistance. This condition is the thread that runs through many chronic afflictions of modern times—obesity, heart disease, and, most conspicuously, type 2 diabetes. All are entangled with diet, and all are linked causally to a dysfunctional response to insulin, the hormone that orchestrates the body's use and storage of nutrients.

Insulin resistance is the fundamental defect in type 2 diabetes, a disease that afflicts 6% of adult Americans, up from 3% in the early 1970s. Most type 2 diabetics are obese, a condition that's so closely associated with insulin resistance that many researchers assume that it is a cause. The prevalence of obesity has increased in the United States almost 2.5-fold since the early 1970s, from 14% to 34%, according to the most recent national surveys.

Metabolic syndrome is another insulin-resistant condition. By some estimates, it afflicts 50 million Americans. It's defined by a cluster of abnormalities—including abdominal obesity, hypertension, and high blood sugar—that precede both coronary heart disease and type 2 diabetes. Stroke, nonalcoholic fatty liver disease, polycystic ovary syndrome, asthma, some cancers, and even Alzheimer's disease have also been associated with insulin resistance.

Once it takes hold, insulin resistance sets up a vicious cycle: As tissues become unresponsive to insulin, the pancreas compensates by secreting ever more insulin, and gradually the tissues grow more resistant.

Elucidating the causes of this destructive cycle is one of the most critical endeavors in modern medicine. Researchers have made progress identifying events that lead to type 2 diabetes and other insulin-involved diseases. But working back up the chain of causality has been a challenge. Unambiguous evidence on the initial stages of disease is missing, making it an excruciatingly difficult task to pin down the causes at the cellular and molecular level.

"The field is in a funny stage right now," says Mitchell Lazar, director of the Institute for Diabetes, Obesity and Metabolism at the University of Pennsylvania. "It's gone from having too few candidate explanations [for insulin resistance] to having too many." Now when someone comes along with yet another possibility, Lazar says, "you go, 'Okay, get in line, buddy.' There are a lot of things that have to be figured out."

Several candidate mechanisms have emerged in the past decade, and two competing theories have gained wide support. One is that cells essentially become poisoned by fat. This lipotoxicity or lipid overload hypothesis holds that normal processes break down when fat (adipose) tissue cannot store excess fat, and fat accumulates inappropriately in muscle and liver cells.

The main rival to this idea, the inflammation hypothesis, is that as fat cells increase in size with the accumulation of fat, they release inflammatory cytokines and molecules known as adipokines. It's these molecules, so this theory goes, that cause insulin resistance elsewhere in the body. Researchers are now confident that these inflammatory mechanisms play some role in insulin resistance. But they still can't say for sure whether those roles are causal.

## Tangled pathways

What makes insulin resistance such an extraordinarily difficult problem to study is that it constitutes "the ultimate systems biology question," says endocrinologist C. Ronald Kahn of the Joslin Diabetes Center in Boston, which is affiliated with Harvard Medical School.

Insulin is the primary regulator of fat, carbohydrate, *and* protein metabolism; it regulates the synthesis of glycogen, the form in which glucose is stored in muscle tissue and the liver, and it inhibits the synthesis of glucose by the liver. It stimulates the synthesis and storage of fats in fat depots and in the liver, and it inhibits the release of that fat. Insulin also stimulates

**Epidemic.** Type 2 diabetes, a disorder of insulin resistance, is on the rise.

the synthesis of proteins and of molecules involved in the function, repair, and growth of cells, and it functions as a signaling molecule conveying information on fuel availability from the periphery to the brain and central nervous system.

"Compared with other hormones," the late J. Denis McGarry of the University of Texas Southwestern Medical Center wrote in *Science* in 1992 (30 October 1992, p. 766), "insulin elicits a bewildering array of metabolic responses in target cells. Deciding which of these are dependent or independent events continues to pose a major challenge." Sixteen years later, that assessment still holds true.

A fundamental role of insulin is to orchestrate the use of fuels in the body, partitioning them to oxidation or storage. When blood sugar is elevated—during and immediately after a meal, for example— insulin works to store excess calories as fat in the fat tissue and transport glucose into muscle tissue. It also signals the mitochondria to use glucose as the primary fuel source. As insulin and blood sugar levels drop in the hours after a meal, they allow fatty acids to be mobilized from stored fat and signal the mitochondria to oxidize these fatty acids for conversion into energy. This "metabolic flexibility," or the capacity to switch easily between glucose and fat for fuel, is a key feature of healthy individuals, as endocrinologist David Kelley, now at Merck Research Laboratories in Rahway, New Jersey, has pointed out.

In insulin resistance, these natural responses break down and become pathological. A "natural system of feedback loops," as Merck's Luciano Rossetti says, is overwhelmed or degraded and disease is often the response. The operative word, though, is "often."

Even among healthy individuals, measurements of insulin-stimulated glucose uptake, insulin sensitivity, and insulin resistance will vary by 600% to 800%. "There's an enormous range," says endocrinologist Gerald Reaven of Stanford University in Palo Alto, California, who deserves much of the credit for persuading the medical research community to take insulin resistance seriously as a causal factor in heart disease and type 2 diabetes. A quarter of this variation in sensitivity can be explained by variations in physical fitness, and another quarter by weight, a relationship that Reaven says has held up in studies of populations as diverse

as the Pima Indians of Southwest Arizona and descendants of Europeans living in Palo Alto. "Clearly, the more obese you are, the more insulin resistant you are," Reaven says, but the same variation can be found in obese subjects, a third of whom are relatively insulin sensitive.

Without being able to pinpoint the tissue, organ, and cell type in which insulin resistance first manifests itself, says Stephen O'Rahilly, co-director of the Institute of Metabolic Science at the University of Cambridge in the United Kingdom, it's virtually impossible "to unpick the causal chain."

What researchers almost invariably measure, though, is how the entire body responds



**Fuel economy.** Among insulin's many functions is as partitioner of metabolic fuels—carbohydrates, fats, and protein— for use and storage in tissues.

to insulin, not how the individual tissues and organs do. And how the body responds also changes dramatically over the course of a day, and from day to day, in response to food intake or physical activity. "We're studying a phenomenon that is happening differentially over a 24-hour period," says O'Rahilly. "But most studies are done when the subject or patient is fasting. Those are essentially looking at insulin's dialog with the liver and how

sensitive the liver is to insulin. It tells you very little about how sensitive the skeletal muscle or the adipocyte is."

### Fat overload

In the mid-1970s, endocrinologists focused on the insulin receptor itself as a likely key to the puzzle. They assumed that resistance was caused either by down-regulation of the insulin receptor—a normal desensitization process—or by a defect in the receptor itself or the binding of insulin to the receptor. By the mid-1980s, Jerrold Olefsky, now at the University of California, San Diego, had demonstrated that the primary defect was downstream in the signaling pathway, not in the receptor itself.

Since the early 1990s, the observation that insulin resistance is associated with elevated levels of free fatty acids in the bloodstream has led researchers to focus on lipid overload as the precipitating event. Several observations support the hypothesis. The single best predictor for the presence of insulin resistance in young, lean offspring of type 2 diabetics, according to Gerald Shulman, an endocrinologist at Yale University, is the accumulation of fat inside muscle cells. Shulman and his colleagues have also studied sedentary populations of lean, healthy, elderly subjects and obese, insulin-resistant adults and children. In all those cases, he says, "the more fat inside the muscle cells, the more insulin resistant they are."

Using nuclear magnetic resonance spectroscopy to do noninvasive measurements of metabolic fluxes—what Shulman calls "basically real-time biochemistry in humans"—Shulman and his colleagues have established that when fat accumulates inside muscle cells, it blocks an intracellular chain of events that normally triggers glucose transport into the cell. The specific culprit, according to Shulman, is the buildup of diacylglycerols (DAGs)—an intermediate product in the formation of triglycerides, the form in which fat is stored in cells. When DAGs accumulate inside muscle cells or liver cells, Shulman has found, they shut down the insulin-signaling pathway. In the muscle cells, they do so by inhibiting the translocation of a protein, glut4, to the cell membrane, where it would normally work to pump glucose into the cell. Insulin-stimulated glucose transport

**Fat as they come.** Researchers made a mouse that can accumulate huge amounts of fat, as the one on the left does by over-expressing adiponectin. The result: This mouse was insulin sensitive.

no longer works efficiently, and the cell is insulin resistant.

"Over the last decade," says Shulman, "we've been able to test this hypothesis using the power of mouse genetics. In more than a dozen transgenic gene knockout models, any time we raise intracellular diaclyglycerols, the mice get insulin resistance in the target tissues; any time we lower it, we prevent insulin resistance." In March, Shulman and his colleagues reported in the journal *Cell Metabolism* that DAG accumulation can also account for insulin resistance in the liver caused by the consumption of high-fructose diets.

The way to think about this, says Shulman, is that the concentration of DAG in a cell is balanced between the delivery of fat to the cells (in the form of fatty acids), the oxidation of fat, and the storage of the fat as triglycerides. "Any time you alter that balance to get a net accumulation of DAGs, through more delivery or decreased oxidation, you get insulin resistance. Anything that flips the balance the other way"—by blocking entry of fatty acids into the cell, for instance, promoting fatty acid oxidation, or even promoting the conversion of DAGs into triglycerides—"prevents insulin resistance." In that sense, the DAGs work as both an intermediate form of a storage molecule and a signaling molecule that tells the cell whether fatty acids are accumulating and whether it's necessary or beneficial to continue pumping in glucose.

**Inflammation**

Competing with the lipid overload hypothesis is the theory that inflammation is to blame. The idea was sparked in the mid-1990s, when Gökhan Hotamisligil of the Harvard School of Public Health and Bruce Spiegelman of Harvard Medical School reported that the inflammatory cytokine TNF-α was overexpressed in animal models of obesity. They demonstrated that they could induce insulin resistance in fat cells in vitro by exposing them to TNF-α. They also showed that they could protect obese strains of mice from insulin resistance by knocking out the genes either for TNF-α itself or for TNF-α receptors.

The hypothesis began to gain wide acceptance after Steven Shoelson of the Joslin Diabetes Center reported in 2001 that he could make cells insulin resistant by overexpressing IKK-β, a molecule that works in signaling pathways to activate the inflammatory mediator NF-κB. Among the compounds that inhibit IKK-β are salicylates, aspirin-like compounds that are used at high doses to treat rheumatoid arthritis and rheumatic fever, both inflammatory conditions. "That struck a chord with me," says Shoelson, because "among the list of things that can cause low blood sugar are salicylates." One obvious implication, he says, is that "inflammation is a potential pathogenic mediator of both insulin resistance and type 2 diabetes."



*"If you can't store fat properly, it's going to build up in liver and muscle and cause insulin resistance."*

—GERALD I. SHULMAN,
HHMI AND YALE
UNIVERSITY SCHOOL
OF MEDICINE

Since then, Shoelson has demonstrated in a series of studies through 2005 that insulin resistance can be induced in lean strains of mice by overexpressing NF-κB in their fat or liver cells and that obese mice can be protected from insulin resistance by inhibiting NF-κB expression. Last year, Shoelson and his colleagues published the results of a pilot study in *Diabetes Care* demonstrating that salicylate therapy could indeed both control blood sugar and reduce inflammatory mediators in obese subjects. Meanwhile, Hotamisligil has linked yet another molecule involved in inflammation, JNK, to obesity and insulin resistance. JNK plays a "predominant role" in the regulation of insulin sensitivity, Hotamisligil wrote on 5 September 2008 in *PloS ONE*: It is overexpressed in animal models of obesity, and knocking it out in these animals both decreases their adiposity and protects them from insulin resistance.

Hotamisligil now believes that the primary cause of JNK activation is stress in the cell's endoplasmic reticulum, which functions to synthesize and fold proteins. In fat tissue, it works to package complex lipids such as cholesterol and triglycerides. Stress in the endoplasmic reticulum will activate JNK, says Hotamisligil, and it's easy to imagine that the demands put on the endoplasmic reticulum by an expanding fat cell are the source of the stress.

The picture that's coming together is that obesity is a low-grade inflammatory state. Excess fat or at least large, overstuffed fat cells activate the immune system, promoting "elevated levels of inflammatory cytokines—IL6, TNF-α, JNK, all kinds of stuff," says Guenther Boden of Temple University in Philadelphia.

A primary source of these inflammatory signals is now believed to be macrophages trapped in the adipose tissue, a discovery first made in 2002 by Anthony Ferrante and his colleagues at Columbia University and, independently, by Hong Chen and colleagues at Millennium Pharmaceuticals. In lean humans or animals, says Ferrante, 5% of the cells in adipose tissue will be macrophages, compared with upward of 50% in obese humans or animals. What recruits the macrophages into the adipose

CREDITS (TOP TO BOTTOM): JA-YOUNG KIM AND PHIL SCHERER/UNIVERSITY OF TEXAS SOUTHWESTERN MEDICAL CENTER, DALLAS; COURTESY OF GERALD I. SHULMAN
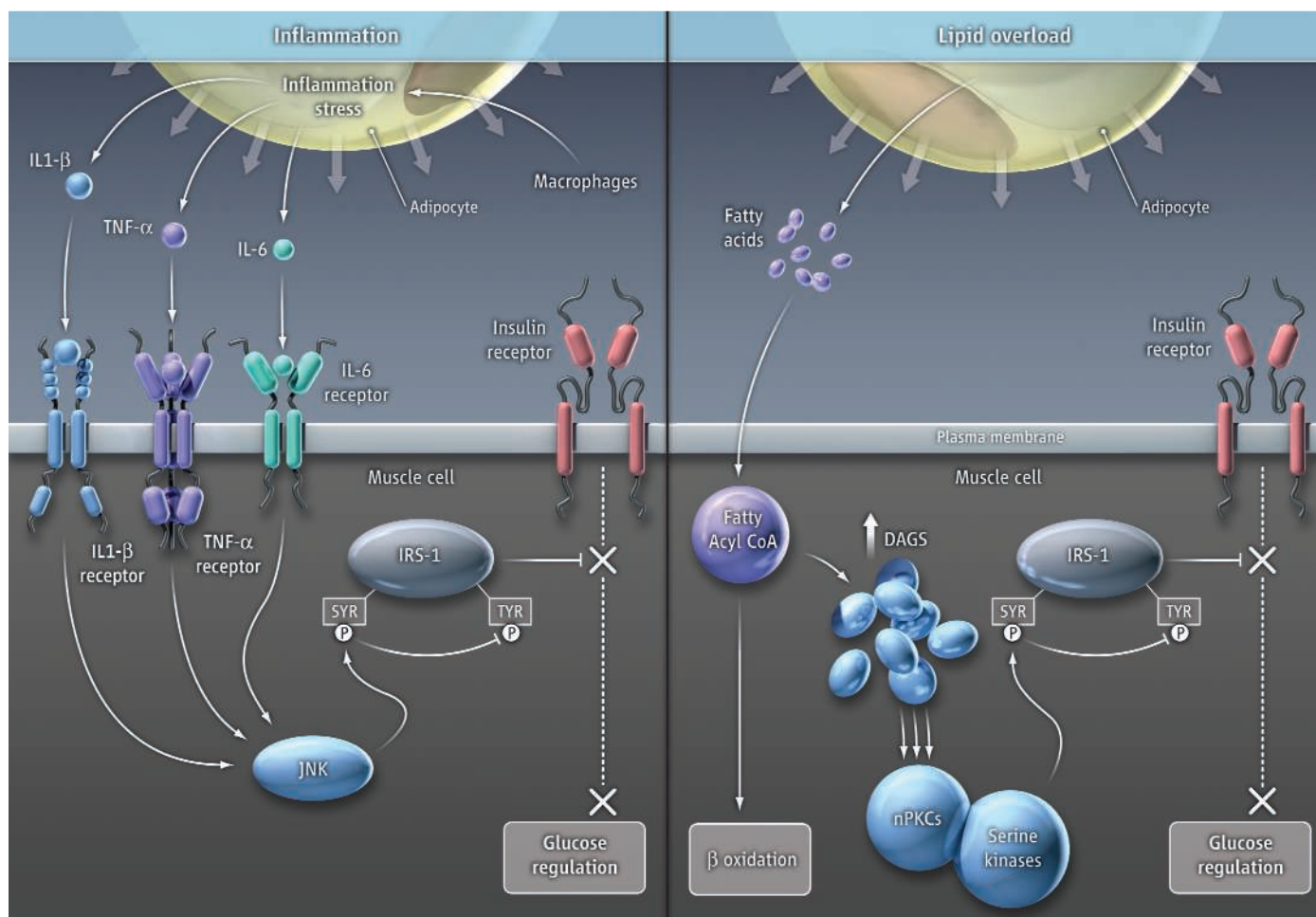
**Main contenders.** Two explanations for the mechanism of insulin resistance have emerged: inflammation (*left*) and lipid overload (*right*). In the inflammation hypothesis, enlarged fat cells attract macrophages and excrete inflammatory signals that work in the muscle cell, via the kinase JNK, to block an insulin receptor substrate (IRS-1) and shut down the insulin-signaling pathway. In the lipid-overload hypothesis, enlarged fat cells leak fatty acids, causing diacylglycerols (DAGS) to accumulate in muscle cells. These inhibit insulin signaling through nPKCs and then block the insulin receptor substrate IRS-1.

tissue is still an open question. Nonetheless, says Hotamisligil, "it's pretty clear that if there are inflammatory cytokines or stress signals around, the insulin receptor does not function very well."

**Focusing on fat tissue**

When researchers discuss their favored hypotheses of insulin resistance, the metaphor that often comes to mind is a Russian nesting doll. Elucidate one mechanism of causation, and it immediately implies the existence of yet another mechanism further down the causal pathway that might be still more fundamental. The end point in this progression, however, invariably seems to be the fat tissue itself.

This has been a consistent theme in insulin-resistance research going back to the early days. Consider, for instance, that impaired glucose uptake by skeletal muscle has traditionally been perceived as the major contributor to insulin resistance. But one rea-

son blood sugar is elevated in type 2 diabetes after a meal, and the primary reason it remains elevated during fasting conditions, is because the liver continues to synthesize glucose and pump it out into the bloodstream even when that glucose is no longer needed.

Insulin was always thought to suppress this process directly, and researchers believed that its failure to do so was a direct manifestation of insulin resistance by liver cells. But Richard Bergman of the University of Southern California in Los Angeles and his colleagues demonstrated in the mid-1990s that this failure to inhibit glucose production in the liver is actually an *indirect* effect of insulin resistance, and that the real location of the insulin resistance is at the fat tissue. What happens, says Bergman, is that the fat cells become resistant to insulin, which then fails to efficiently suppress, as it should, the release of fatty acids from these cells. It's those liberated fatty acids that in turn stimulate the inappropriate production

of glucose by the liver. "We believe most of the effect of insulin on the liver is indirect," Bergman says, "and it's mediated by free fatty acids" released from the fat tissue. (Complicating matters further, some of the apparent failure of insulin to suppress glucose production in the liver, as Rossetti and collaborators have demonstrated, also appears to be mediated by insulin resistance in the brain.)

Most researchers now believe that both inflammation and DAG accumulation are causal factors in insulin resistance, but this raises the obvious question of what causes the inflammation, and what causes the DAGs to accumulate in muscle and liver cells in the first place. One likely possibility, says Shulman, is that people simply eat too much for their level of physical activity. The excess nutrients, in this scenario, then overwhelm the fat tissue, causing the fat cells to expand and secrete inflammatory molecules, or they spill out of the fat tissue and

instead accumulate where they don't belong. But that doesn't explain why some obese individuals—often very obese—remain resolutely insulin sensitive. This suggests that something about the fat tissue itself, and maybe its ability to absorb and retain fatty acids and do so in a manner that doesn't induce inflammation, is the fundamental defect, the critical factor determining whether fatty acids will accumulate as triglycerides in healthy fat depots or as DAGs in liver and muscle cells.

A telling piece of evidence, suggests Shulman, is that insulin resistance is also common in rare genetic disorders known as lipodystrophies, which are characterized by a deficiency or complete absence of fat cells. Lipodystrophic individuals have little or no place to temporarily store the calories they consume before they use them for fuel, and they are extremely insulin resistant. Researchers have also created lipodystrophic mouse models, genetically manipulated to have no fat cells, and these are also extremely insulin resistant. The fact that insulin resistance occurs in mice and humans lacking the fat cells necessary to store excess nutrients, says Shulman, "suggests that if you can't store fat properly, it's going to build up in liver and muscle and cause insulin resistance."

So what does it mean to store fat properly? The key, some researchers say, is the ability to expand adipose tissue in a specific way. When fat tissue can generate new adipocytes, these researchers believe, it creates fresh storage capacity instead of shunting excess fat into existing, overstuffed fat cells. According to this hypothesis, insulin resistance develops when fat cells are overstuffed, stressing the endoplasmic reticulum and attracting macrophages, releasing inflammatory mediators, or leaking fatty acids out into the circulation—or any combination of these.

Among the evidence supporting this hypothesis is a transgenic mouse created by Philipp Scherer of the University of Texas Southwestern Medical Center at Dallas and his colleagues in 2007. It happens to be, as Scherer says, "probably the fattest mouse ever made." It's also extremely insulin sensitive. This particular mouse overexpresses a molecule called adiponectin, discovered by Scherer in 1995, that appears to stimulate the formation of new fat cells. Scherer says his transgenic mouse continues to generate new small fat cells which can "deposit all these calories taken in into an expandable healthy fat pad." The liver stays in pristine shape, he says: "There's no lipid

accumulation [even in muscle cells], … and there's improved insulin sensitivity."

Another line of evidence supporting this hypothesis comes from experience with insulin-sensitizing drugs, known as thiazolidinediones, used to treat type 2 diabetics. These drugs target a receptor on cells, called PPARγ, that also works in the subcutaneous fat tissue to differentiate new adipocytes. The diabetic patient gets fatter, but the excess is stored in new small fat cells rather than in overstuffed old ones. The patient gains insulin sensitivity as a tradeoff for the extra fat. "You redistribute fat out of the muscle, liver, and beta cells into subcutaneous fat," says Ralph DeFronzo, chief of the diabetes division at the University of Texas Health Science Center at San Antonio. "As long as the fat is in subcutaneous adipocytes, it can't hurt you."



"It's pretty clear that if there are inflammatory cytokines or stress signals around, the insulin receptor does not function very well."
—GÖKHAN. S. HOTAMISLIGIL, HARVARD SCHOOL OF PUBLIC HEALTH

### Good and bad fat cells?

While researchers have made considerable progress elucidating these mechanistic connections, every insight seems to come with unanswered questions or observations that remain stubbornly controversial.

Take the critical observation that fatty acid levels are elevated in obesity, and the idea that this leads to DAG accumulation in liver and muscle cells. At least some researchers—Keith Frayn, for instance,

who studies adipose tissue metabolism at the University of Oxford in the United Kingdom—question whether this is true. "Every review of insulin resistance talks about an increase in free fatty acids" with obesity, Frayn says. "We have been looking at a collection of 1200 normal healthy individuals, and we see no correlation in that collection between body mass index and free fatty acids in plasma."

The evidence that large, overstuffed fat cells are the problem has also recently been challenged. Reaven and Sam Cushman, a fat metabolism researcher at the U.S. National Institute of Diabetes and Digestive and Kidney Diseases in Bethesda, Maryland, reported in August 2007 that when they look at subjects with the same level of moderate obesity but different degrees of insulin sensitivity, they find that the insulin-resistant subjects actually have fat cells that tend to be smaller, rather than larger.

"The conventional wisdom has been that the obese have these very big fat cells, and these secrete all these terrible things [inflammatory cytokines in particular], and these terrible things make you insulin resistant," says Reaven. "What we found is that if you looked at the ratio of small fat cells to large, insulin-resistant people had the higher proportion of small cells." To Reaven, this suggests that the underlying problem in insulin resistance isn't the large fat cells themselves but a relative inability to expand smaller fat cells into larger ones as needed. "If you can't make good fat cells to store fat," he says, "then the fat may end up in ectopic places where it does more harm than good."

One observation that seems indisputable is that when individuals lose weight, they become more insulin sensitive. If nothing else, this has given researchers the confidence to assume that excess body fat—particularly in the abdomen and around the internal organs—is a fundamental cause of insulin resistance. But that still avoids the question of what causes insulin resistance in lean individuals. This is something few researchers will even address, although one possibility is that they, too, simply can't store fat safely in subcutaneous pads.

"The biggest question in the whole field of insulin resistance is still this direction of causality," says O'Rahilly. "Does obesity make you insulin resistant? Or does underlying factor x cause both obesity and insulin resistance?"

—GARY TAUBES

**Looking ahead.** The Senate is expected to approve Charles Bolden and Lori Garver to lead NASA.

## U.S. SPACE PROGRAM

# Can Bolden Banish NASA Blues?

**Obama's nominee tries to rally public support on eve of moon landing anniversary as an expert panel and Congress worry that NASA has lost its way**

Next week marks 40 years since NASA first put men on the moon. But trepidation about the space agency's future is dampening celebration of that milestone achievement. "NASA is not what it was," declared the chair of the Senate Commerce, Science, and Transportation committee during last week's confirmation hearing for Charles Bolden Jr., President Barack Obama's choice to lead the agency.

Rather than inspiring the nation with wondrous feats of science and exploration, said Senator Jay Rockefeller (D–WV), NASA is adrift. The agency's biggest fan on Capitol Hill agrees. "That magic is gone," bemoaned Senator Bill Nelson (D–FL), who once flew on the soon-to-be-retired shuttle. Even Bolden, a 62-year-old retired Marine general and astronaut, joined the chorus of gloom during a hearing in which legislators lauded his fitness for the job.

But some remedies to this midlife crisis are emerging. The day before Bolden testified, the National Academies' National Research Council (NRC) released a report urging the space agency to link its efforts to broader national goals. Both the report and Bolden suggest that NASA support more basic research, as it once did, and take the lead in monitoring the environment. Both thrusts will require greater cooperation with other nations, they add.

The biggest question facing NASA is whether the 2004 vision of President George W. Bush to return humans to the moon by 2020 and then on to Mars is still alive. During the presidential campaign,

Obama promised to build a large new rocket that can put humans back on the lunar surface, and Bolden told senators that is still in the cards. "We will go on to the moon," he said. But he avoided mention of any timeline and left up in the air the second phase of Bush's vision. "I want to go to Mars," Bolden declared, before noting that this would be at least a 20-year venture.

Next month, a blue-ribbon panel chaired by retired aerospace executive Norman Augustine will lay out options for the replacement of the shuttle and whether it will be designed with a lunar base and a Mars mission in mind. But the more salient issue is how the White House will react.

The academies' report makes some suggestions. *America's Future in Space: Aligning the Civil Space Program with National Needs* argues that NASA needs shaking up so that "a disciplined space program can serve larger national imperatives." Toward that end, the panel recommends that NASA create a nimble research shop modeled on the Defense Department's Defense Advanced Research Projects Agency.

In his Senate testimony, Bolden strongly backed a renewed emphasis on basic science and engineering, particularly aeronautics. His nominated deputy, Lori Garver, told legislators that she foresees "a great future in utilizing the space station for biomedical research" once the orbiting facility is completed. She said the research would focus on human diseases, an area largely ignored by the Bush

Administration, and on preparing astronauts for long stints in space.

The academy report and Bolden also agree that NASA must revitalize its Earth-observation system and work with other countries to gather data on the planet's environmental health. Strengthening international ties is essential on other fronts, too, NASA science chief Edward Weiler warned researchers meeting last week in Washington, D.C.: "On our own, we can't do what people would like us to."

Mars is a case in point. "We no longer have a viable Mars program," Weiler confessed to a planetary science decadal study group. To help build one, NASA and the European Space Agency tentatively agreed early this month at a meeting in Plymouth, U.K., on a cooperative Mars robotic exploration effort. Although details from that meeting have yet to be made public, one NASA official said that finding life on the Red Planet is high on the list.

The NRC study also urged NASA to expand its roster of partners in human exploration beyond Europe, Japan, Canada, and Russia. Neither Bolden nor the White House has spoken publicly about this idea, but several Washington officials predicted that Obama's efforts to strengthen relationships with other countries, notably China and India, will eventually include space cooperation.

The former shuttle pilot also expects help from the private sector. "The government cannot fund everything we want to do," he said, calling for entrepreneurs to take a larger role.

Bolden, who grew up in South Carolina during segregation and who would be the first African American to lead NASA, faced no tough questions during the hearing. Nelson, who accompanied Bolden on a 1985 shuttle mission, called him an "overcomer" of personal, racial, and professional barriers. Concerns about his role as an aerospace consultant appear to have dissipated.

The Senate is expected to confirm Bolden easily although not necessarily before it goes on holiday in August. Once he takes the helm, Bolden's biggest challenge will be to win support from the White House for a new approach to space exploration that fires the imagination of Congress and the public. If he fails, the agency may be left celebrating past triumphs rather than working toward future milestones.

**–ANDREW LAWLER**

With reporting by Richard A. Kerr.

# Online
**sciencemag.org/ apollo11**

Classic papers from the Apollo 11 mission.

**Hard times.** A catastrophic rearrangement of the outer planets may have pummeled Earth's moon with huge impacts.
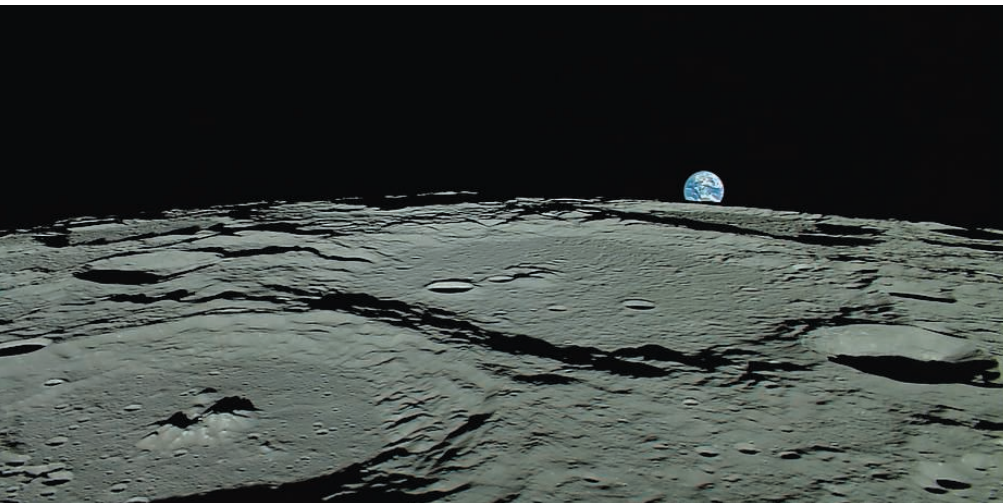
SOLAR SYSTEM EVOLUTION

# Shifting Orbits Gave Solar System A Big Shakeup, Model Suggests

**Dynamicists simulating the solar system's early days are finding that a violent reshuffling of bodies large and small may explain many of today's planetary mysteries**

Isaac Newton inspired generations with his vision of the planets, their moons, and all the miscellaneous flotsam of the solar system as one huge planetary clock that has kept precise time over the eons. But planetary scientists are finding that the most massive pendulums within the planetary timepiece—the four outermost planets—haven't always been orbiting where they are today. They've moved, some a considerable distance outward.

The most catastrophic scenario for such planet migration, dubbed the Nice model (after the French city), has been gaining ground of late. It envisions the great reshuffling as a brief, violent affair that not only put the outer planets where they are today but also created the Kuiper belt of small icy bodies beyond Neptune, gave the planets scores of oddly orbiting moons, and bombarded the solar system with a rain of asteroids and comets so fierce that it would have cooked all but the deepest subterranean life on early Earth.

If such a cataclysmic rearrangement did indeed occur, "almost every nook and cranny of planetary science has been affected by it," says planetary dynamicist William Bottke of the Southwest Research Institute (SwRI) in Boulder, Colorado. The latest support for the Nice model, a new explanation for primitive-looking asteroids, appears this week in *Nature*. But the model has more hurdles to clear, such as explaining why the innermost planets—Earth and its neighbors—weren't reshuffled as well.

### A French quartet

Today, the sun's planets fall tidily into two neighborhoods: the inner solar system, home of small, rocky Mercury, Venus, Earth, and Mars; and the outer solar system, where enormous Jupiter and its smaller (but still huge) gaseous cousins Saturn, Uranus, and Neptune trace out stately decades-long orbits. In between lies a no planet's land of rubble, the asteroid belt.

About 10 years ago, planetary dynamicists realized that the outer planets had moved over time and that migrating planets could dump small bodies such as comets and asteroids into the inner solar system like salt from a shaker. But there was a problem: The solar system had formed 4.6 billion years ago, but astronomers saw no sign of such a cosmic bombardment until 700 million years later. How could trillions upon trillions of bodies have been kept locked up in the outer solar system for so long, only to be suddenly unleashed?

Four planetary dynamicists had an idea (*Science*, 3 December 2004, p. 1676). Harold Levison of SwRI, Rodney Gomes of the National Observatory in Rio de Janeiro, Brazil, and Kleomenis Tsiganis of Aristotle University of Thessaloniki, Greece, had all at one time or another conducted research with astronomer Alessandro Morbidelli at the Observatory of the Côte d'Azur in Nice. They began with a scheme in which the outer planets had formed much closer to the sun than they are now. Immersed in a disk of planet-building debris, however, the planets wouldn't have stayed where they formed. Any time a planet encountered a planetesimal and gravitationally flung it away, the planet would drift an infinitesimal amount in reaction. Massive Jupiter barely budged, but Saturn and the other relative lightweights crept inexorably away from the sun.

Eventually, Saturn's outward drift would have brought Jupiter and Saturn into their so-called 1:2 resonance, in which Saturn made two orbits in the time it took Jupiter to make one. (Today, the ratio is closer to 1:2.5.) Then Jupiter could repeatedly give Saturn a gravitational nudge at the same point in Saturn's orbit so that the nudges could accumulate, the way repeatedly pushing a swing at the same point in its arc sends it higher.

The repeated orbital boosts would have stirred the outer solar system into a frenzy. The Nice group ran computer simulations of the gravitational interactions of the planets and planetesimals with the planets starting bunched in close. Once locked in their resonance, the model's Jupiter excites the orbit of Saturn, stretching its orbit until it can gravitationally scatter the much smaller Uranus and Neptune outward into a lingering disk of planetesimals. In all the hubbub, the two outermost planets can even cross orbits and exchange positions.

The model's Neptune then sends planetesimals every which way. A million billion of them pummel Earth's moon in less than 100 million years; the inner planets fare no better. "This is a very violent event," Levison said at a workshop last November.* "The solar system rearranges itself, and the inner solar system gets clobbered."

In three papers published in 2005, the Nice group reported signs that such a rearrangement and bombardment may have actually happened. The slightly tilted and elongated final orbits of their model's outer planets resemble the planets' actual orbits, they reported. All earlier modeling had left the outer planets with perfectly circular orbits with no tilt.

### Planetary forensics

Encouraged, the Nice group began looking

CREDIT: JAXA

farther afield for clues to the early solar system's evolution. "It's like a bloody crime scene," says Levison. "Sometimes the splatters on the wall can tell you more about the crime than the body itself." First, the Nice group looked at Jupiter's Trojan asteroids, the small bodies that lead and trail the giant planet approximately in its own orbit. According to a leading theory, they got there when Jupiter's growing envelope of gas dragged planetesimals into the planet's orbit, but that theory never explained why some of the Trojans move in steeply inclined orbits.

In Nice model simulations, when Jupiter and Saturn cross their 1:2 resonance, the Trojans naturally turn up just as they should, the group reported in one of its 2005 papers. The modeled Trojans' range of orbits matches the actual orbits "remarkably" well, they wrote. The final number of Trojans in the model fit the observed number. And the model has them coming from the same reservoir of icy planetesimals as the comets, which would explain the Trojans' cometlike spectral color.

The Nice group also hypothesized that similar resonances between migrating Uranus and Neptune could produce the Trojans in Neptune's orbit. Planetary dynamicists David Nesvorný of SwRI (and a some-time collaborator of Nice group researchers) and David Vokrouhlický of Charles University in Prague confirmed that prediction this June in *The Astronomical Journal*.

Since 2005, Nice group members and colleagues have found more fingerprints of a sudden and violent planetary rearrangement all over the solar system, from the asteroid belt to the solar system's outer limits. The four outer planets have distant moons whose origins and behavior have been hard to explain. These moons wing about every which way, half of them revolving "backward." Unlike the inner, well-behaved moons, which obviously formed like mini–solar systems from disks of debris, the more distant, irregular satellites must have arrived from elsewhere, but how they managed to get into orbit remained a mystery.

In a 2007 paper in *The Astronomical Journal*, Nesvorný, Vokrouhlický, and Morbidelli showed how close encounters of two planets—like those in the Nice model's migration scenarios—can deflect a nearby planetesimal into orbit around a planet. The modeled orbits and the number of captured moons compared well with observations for moons around Saturn, Uranus, and Neptune.

"That's another success," says Levison.

Beyond the outer planets, Levison and Nice colleagues used their model to try to understand the Kuiper belt, the disk of icy bodies orbiting the sun beyond Neptune. They were "able to explain the basic mysterious aspects of the [Kuiper belt] population," says Levison—no mean feat in a region where oddities abound. For example, Kuiper belt objects orbit inside the 1:2 resonance with Neptune, as if it presented a barrier. Some belt objects travel in resonance with Neptune, and some do not. Some orbits are inclined, and some are not. And some Kuiper belt bodies are sorted into high- or low-inclination orbits depending
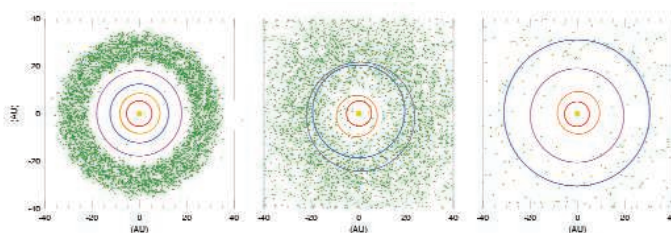




**A big kablooie.** When the outer planets (colored orbits) rearranged themselves (*center*), they scattered planetesimals (green dots), including Saturn's 213-kilometer irregular satellite Phoebe (*above*).

on their size and color. In Nice model simulations that Levison and colleagues reported in 2008 in *Icarus*, seven of these peculiarities appeared naturally in the Kuiper belt. "It looks pretty good," says Morbidelli. "Most features are explained within one scenario."

Levison and colleagues report this week in *Nature* that the dark, organic-rich residents of the outer asteroid belt could be icy interlopers thrown there by migrating planets. Most theorists have assumed that the outer belt's water-rich D-type asteroids formed where they are now, at relatively low temperatures, whereas the higher temperatures of the inner belt baked

out any water and organics, leaving rocky bodies. But in Nice model runs, Jupiter slings icy planetesimals inward where jovian resonances capture them, tame their gyrating orbits, and then move on, leaving the newly arrived "asteroid" orbiting toward the outer parts of the asteroid belt.

## Growing acceptance

"It's scary," says Levison. "It's a crazy idea, and it's working remarkably." He's particularly impressed with the way the Nice model creates gravitational conditions essential to producing different aspects of the solar system, such as the Trojans and irregular satellites. No alternative explanation—such as the slow migration of planets without a resonance crossing—has those essential conditions, Levison says.

Many other planetary scientists are impressed, too. "The overall model has held up quite well," says planetary dynamicist Stuart Weidenschilling of the Planetary Science Institute in Tucson, Arizona. "It's pretty much accepted as the paradigm for how the solar system could have evolved." And researchers in and out of the Nice group are exploring for possible implications elsewhere: the inner and outer Oort cloud of comets far beyond the Kuiper belt; the interiors of Jupiter's large satellites; and the mysterious dark stuff coating some satellites of the outer planets, among other places.

Although the Nice model may be on a tear, "there's still some more work to be done," says Weidenschilling, to prove that it really happened that way. For one thing, there's a problem with Earth and its fellow inner planets. "The inner planets are not stable in the Nice model," notes planetary dynamicist Renu Malhotra of the University of Arizona, Tucson. For example, in the model, Mars might fly out of the solar system as resonances of migrating outer planets sweep through.

And delaying the heavy bombardment to 3.9 billion years ago is tricky. Even with the 1:2 resonance of Jupiter and Saturn to light the fuse, it requires that material in a belt beyond the newly formed outer planets "delicately hang around for 700 million years, essentially doing nothing," says Malhotra. "That seems very difficult." The Nice group is working on both problems and is guardedly optimistic about solving them. "We haven't got much of an alternative," says Morbidelli. "The only coherent scenario is ours."

**–RICHARD A. KERR**

CREDITS (TOP TO BOTTOM): NASA/JPL; HAROLD LEVISON/SWRI

# Switch to Superior High Fidelity PCR

# PrimeSTAR®

## Top Five Reasons to Switch to PrimeSTAR®:

▶ **Higher Accuracy:** A strong exonuclease activity results in an extremely low error rate, with only 15 errors per 480,000 bp on a GC-rich template.

▶ **Higher Efficiency:** Higher than *Taq* Polymerase.

▶ **Robust Amplifications:** A single PCR cycling protocol can be used to amplify products of varying sizes.

▶ **Greater Specificity:** The antibody mediated Hot Start formulation prevents false initiation events during reaction assembly.

▶ **Excellent with GC-Rich Targets:** Robust performance with GC-rich templates using the GC buffer formulation.

**Amplification Efficiency of a 2 kb Human Genomic DNA Fragment** (DCLRE1A)**.** Excellent sensitivity and yield are observed when amplifying with PrimeSTAR®. Quantities: Lane1: 0 ng (dH$_2$O), Lane 2: 100 pg, Lane 3: 1 ng, Lane 4: 10 ng, Lane 5: 100 ng.



PrimeSTAR® | M 1 2 3 4 5 — Company N 1 2 3 4 5 — Company B 1 2 3 4 5 — Company I 1 2 3 4 5 — Company S 1 2 3 4 5

## TaKaRa

For more information
**www.takara-bio.com**

**Japan:**
Takara Bio Inc.
+81 77 543 7247
www.takara-bio.com

**USA:**
Takara Bio USA
A Division of Clontech Laboratories, Inc
888-251-6618
www.takarabiousa.com

**Europe:**
Takara Bio Europe S.A.S.
+33 1 3904 6880
www.takara-bio.eu

**China:**
Takara Biotechnology (Dalian) Co., Ltd.
+86 411 8764 1681
www.takara.com.cn

**Korea:**
Takara Korea Biomedical Inc.
+82 2 2081 2525
www.takara.co.kr

# LETTERS

edited by Jennifer Sills

## Making the Most of Online Collaboration

IN HER PERSPECTIVE ("TECHNOLOGY AND INFORMAL EDUCATION: What is taught, what is learned," 2 January, p. 69), P. M. Greenfield emphasizes that scientific thinking should go beyond mere visual literacy to include "reflection, analysis, critical thinking, mindfulness, and imagination." Informal learning on the Internet can encompass facets other than visual media. As avid readers and writers of science at the online international undergraduate *The Journal of Young Investigators* (JYI) (*1*), we believe that informal education through Internet-based media promotes active and critical scientific thought.

Despite the prevalence of visual media, text remains the primary form of communication and collaboration on the Internet. Online science media such as JYI provide opportunities for students to critically read and write about literature by constructing logical scientific arguments. Practicing scientists often note that 50% of the scientific process involves strong written communication in the form of grant writing and journal articles. Grant proposals in particular require students to communicate their ideas compellingly to generalists and situate the work in a broader context.

Undergraduate publication experience also introduces students to the world of peer review, a cornerstone of scientific collaboration. They are forced to address criticism of their own research experiments or of their interpretation of the literature. The global reach of the Internet ensures exposure to many schools of thought. JYI itself has students from 60 undergraduate institutions across 14 countries.

Such experiences give students an edge as budding scientists by not only developing strong scientific writing and thinking, but also by encouraging the habit of staying abreast of the latest advances in science outside of the student's area of specialization. Students gain a broad understanding of science's relationship to society, so that they may advocate on science's behalf.

In this way, informal online collaboration involving writing stimulates critical discussion of scientific issues and allows students to actively shape their own education.

**JUSTIN CHAKMA[1]\* AND BETTY PANG[2]**

[1]McLaughlin-Rotman Centre for Global Health, University Health Network and University of Toronto, Toronto, ON M5G 1L7, Canada. [2]Department of Cellular and Molecular Biology, University of Michigan, Ann Arbor, MI 48104, USA

\*To whom correspondence should be addressed. E-mail: justin.chakma@utoronto.ca

### References and Notes
1. *The Journal of Young Investigators* (www.jyi.org).
2. *Science* has provided editorial support to JYI.

## Cognitive Aging Data Will Take Time

IN A PERSPECTIVE ABOUT NONVERBAL IQ ("Technology and informal education: What is taught, what is learned," 2 January, p. 69), P. M. Greenfield appeals to the Flynn effect, which describes an increase in global IQ over the past 100 years, to show that cognitive aging decreased between 1942 and 1992. Greenfield alleges that there is now less loss of nonverbal IQ with aging than there was in the 1940s. However, the data she supplies can neither support nor disconfirm the hypothesis that there has been a reduction in cognitive aging between 1942 and 1992.

The problem is that each age group tested in 1992 represents a different cohort, not the same cohort at different ages. To test her claim about cognitive aging, one would need to compare the 25-year-olds tested in 1942 with the 75-year-olds tested in 1992. Such data are not available. A close proxy is a comparison of the 25-year-olds of 1942 with the 65-year-olds of 1992. Such a comparison suggests that there was no cognitive aging for that cohort. Hence, there may not have been a change in cognitive aging. We will be able to determine this only when we test the 1992 25-year-old cohort when they are 65 years old. **K. BRAD WRAY**

Department of Philosophy, State University of New York, Oswego, Oswego, NY 13126, USA. E-mail: kwray@oswego.edu

### Response

WRAY MAKES A VALID POINT: BECAUSE THE data in Figure 2 are cross-sectional (that is, all ages were given the Raven's nonverbal IQ test in the same year, either 1942 or 1992), there is a confound between age and cohort (that is, at each time of testing, each age increase of 10 years also represents cohorts that were born 10 years earlier in time). Because of this confound, the Figure 2 data are compatible with either of two interpretations (or a combination of both): They are compatible with (i) a decrease in cognitive aging between 1942 and 1992, and (ii) a decrease in the Flynn effect over the decades, with earlier cohorts (people born earlier) showing bigger differences in performance between 1942 and 1992 than later cohorts.

I agree with Wray that only two longitudinal cohorts born in two different historical periods can definitively answer the question of whether cognitive aging has declined. My goal in introducing the idea of a historical decrease in cognitive aging in my Perspective was to

CREDIT: ISTOCK

stimulate systematic empirical investigation explicitly designed to test this hypothesis.

**PATRICIA M. GREENFIELD**

Department of Psychology, University of California, Los Angeles, CA 90095, USA. E-mail: greenfield@psych.ucla.edu

## Open Access: Increased Citations Not Guaranteed

IN THEIR BREVIA "OPEN ACCESS AND GLOBAL participation in science" (20 February, p. 1025), J. A. Evans and J. Reimer report a small but significant citation effect (about 8%) that they attribute to free access to the scientific literature. However, Evans and Reimer only measure the effect of open access where publishing is concerned, such as when a journal makes articles freely available after a period of delay (1). They ignore other sources of open-access articles, such as when authors pay to make their articles freely available in subscription-access journals (2) or use self-archiving. In a randomized controlled trial of open-access publishing, we were unable to detect a citation advantage that could be attributed to access status, although we did observe that open-access articles received more article downloads from more visitors (3).

**PHILIP M. DAVIS**

Department of Communication, Cornell University, Ithaca, NY 14853, USA. E-mail: pmd8@cornell.edu

### References

1. HighWire Press (http://highwire.stanford.edu/lists/freeart.dtl).
2. P. M. Davis, *J. Am. Soc. Inf. Sci. Technol.* **60**, 3 (2009).
3. P. M. Davis *et al.*, *BMJ* **337**, a586 (2008).

## Open Access: The Self-Selection Effect

IN THE BREVIA "OPEN ACCESS AND GLOBAL participation in science" (20 February, p. 1025), J. A. Evans and J. Reimer claim that open access—i.e., free and unrestricted online access to scientific publications—has little influence on research attention, as measured by article citation frequency. Their claim is questionable, however, because it assumes that open access is a randomly assigned journal attribute, whereas it is actually assigned by publishers according to their objectives and the characteristics of the journal.

Large, established, widely distributed journals naturally attract important papers and are, consequently, highly cited. Converting such journals to open access will likely cause a fall in revenue unmatched by a comparable rise in impact, making conversion an unappealing option. Publishers of new stand-alone journals face a different situation. Unless they have a captive market, such as a learned society, they will likely have difficulty selling subscriptions. In such cases, open access appears to offer the best hope for gaining both visibility and a stream of contributions.

Although the Evans and Reimer study indicates little about the influence of open access on the impact of otherwise similar journals, it does establish that, with open access, new journals can be as effective as the old in gaining readership for the work that they publish. This means that established journals have no inherent monopoly over the literature and that the creation of effective new options for distributing research findings remains possible.

**ALFRED N. BURDETT**

Heron Publishing, 202-3994 Shelbourne Street, Victoria, BC V8N 3E2, Canada. E-mail: alfredburdett@heronpublishing.com

## Open Access: The Sooner the Better

IN THE BREVIA "OPEN ACCESS AND GLOBAL participation in science" (20 February, p. 1025), J. A. Evans and J. Reimer argue that a research article published online is only modestly (8%) more likely to be cited if it is freely available. This result would seem to cast doubt on one important argument in favor of free access—that it will increase the visibility of a paper to colleagues.

However, the 8% statistic that Evans and Reimer highlight is misleading. The authors' supporting online material (figure S1C) clearly shows that the impact of free access on citations is heavily dependent on the age of the article at the time free access was provided. In particular, when articles were made freely available within 2 years of publication, their citations increased by almost 20%.

This far more dramatic effect is the one scientists and journals should consider when deciding when to provide free access. If this decision is to be made purely on the basis of citation impact, the upward trend of the curve in figure S1C argues strongly in favor of minimal delays.

Unfortunately, it is hard to tell exactly how short a delay the data support, because the underlying citation information is not provided. That the raw data for such a provocative paper is unavailable is an astonishing violation of the norms of science, and the explicitly stated publication policies of *Science*.

**MICHAEL EISEN[1]\* AND STEVEN SALZBERG[2]**

[1]Howard Hughes Medical Institute, Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA. [2]Center for Bioinformatics and Computational Biology, Department of Computer Science, University of Maryland, MD 20742, USA.

*To whom correspondence should be addressed. E-mail: mbeisen@berkeley.edu

*Editor's Note: It is* Science *policy, as stated in our online information to contributors, that "After publication, all data necessary to understand, assess, and extend the conclusions of the manuscript must be available to any reader of* Science.*" However, we do not preclude our authors from obtaining data from commercial sources when those are the only sources of the data and when those data are available to the scientific community.*

**BRUCE ALBERTS**
Editor-in-Chief

## Response

DAVIS ACCURATELY OBSERVES THAT OUR analysis of open access only accounts for scientific literature provided by publishers. We chose to analyze when journal volumes come online through publisher Web sites because publishers do not select specific articles for availability; instead, they select a batch of articles (one or more years' worth) based on duration since publication. The difficulty with analyzing the open-access effect for articles that authors have paid or taken pains to post freely is that authors likely select the best articles. Moreover, in author-pays and self-archiving analyses, articles cannot be compared with themselves over time—they must be compared with other articles that were not selected. As a result, they report large open-access effects—from 100% (1) to 286% (2). A reanalysis of one of these studies using instrumental variables to predict whether authors paid the open-access fee suggests that much of the purported open-access effect comes from author selection (3).

Davis recommends his recent open-access experiment with 11 physiology journals that finds no open-access effect (4). Truly randomized, article-level experiments would offer an improvement over current studies. The problem with Davis's experiment, however, is that it introduces another level of selection suggested by Burdett's letter:

Established private journals like *Nature* or *Cell* would never consider opening their content for an open-access experiment. Journals that do are often sponsored by scientific societies and post a much smaller sticker price. Consider that the average price per article in Davis' sample of journals is $3.39 relative to $14.55 for all 66 physiology titles indexed by Thomson's *Web of Science* and includes the eight least expensive journals (the average price in our open access sample was $10.28). It should not be surprising that the cheapest journals post an indiscernible open-access effect.

This illustration validates Burdett's criticism of our analysis: The modest open-access effect we report derives from the subset of journal volumes that are at some but not all points in the public domain by 2004. The effect would likely be larger if the most expensive journals from private publishers had made their holdings available freely.

Eisen argues that the 8% open-access effect we report is misleading because he interprets our figure S1C to suggest that the effect is larger in recent years. A methodological challenge described in our supporting online material cautions against this interpretation. Our analysis relies on estimates of what citations would have been in the absence of online access. Article citations typically trace a log-normal distribution, with a steep rise in citations followed by a gradual fall (*5*). Whether one models this path explicitly, as we do, or simply uses the prior year's citations, as we show in our supporting material, the estimates become less accurate as you approach the present. For very recent years, these calculations underestimate expected citations because this is when the citation trend rises most steeply. This produces an inflated estimate of the influence of free and commercial online availability, exacerbated because journals that become open access do so disproportionately in the last years of our study.

Burdett suggests that new journals can gain quick access to the market for ideas through an open-access model. Our published analysis could not directly support this claim—our estimation excluded journals online at publication—but additional models available from the author provide strong support for it. The only reasonable explanation is that, following Eisen, the culture of modern science and scholarship values the open-access ideal (*6*).

The irony is not lost on us that we published a paper about open access whose data is not open access. It was collected and is owned by private companies. It is, however, widely available and licensed to thousands of research institutions internationally for those who would reassemble it and improve upon our analyses.

**JAMES A. EVANS**

Department of Sociology, University of Chicago, Chicago, IL 60637, USA. E-mail: jevans@uchicago.edu

**References**
1. G. Eysenbach, *PLoS Biol.* **4**, e157 (2006).
2. S. Lawrence, *Nature* **411**, 521 (2001).
3. P. Gaule, N. Maystre, *CEMI-WorkingPaper* (École Polytechnique Fédérale de Lausanne, Lausanne, 2008).
4. P. M. Davis, B. V. Lewenstein, D. H. Simon, J. G. Booth, M. J. Connolly, *BMJ* **337**, a568 (2008).
5. M. J. Stringer, M. Sales-Pardo, L. A. Nunes Amaral, *PLoS ONE* **3**, e1683 (2008).
6. R. K. Merton, *J. Legal Political Soc.* **1**, 115 (1942).

## TECHNICAL COMMENT ABSTRACTS

### COMMENT ON "Neodymium-142 Evidence for Hadean Mafic Crust"

**Rasmus Andreasen and Mukul Sharma**

O'Neil *et al*. (Reports, 26 September 2008, p. 1828) presented neodymium-142 data for rocks from northern Quebec, Canada, and suggested that these rocks may represent the oldest preserved crustal section on Earth. We argue that the age of the rocks is based on a spurious correlation between rocks that are probably not co-genetic and negative $^{142}$Nd anomalies that may be the result of an analytical artifact.

Full text at www.sciencemag.org/cgi/content/full/325/5938/267-a

### RESPONSE TO COMMENT ON "Neodymium-142 Evidence for Hadean Mafic Crust"

**Jonathan O'Neil, Richard W. Carlson, Don Francis, Ross K. Stevenson**

Andreasen and Sharma raise concerns about the neodymium-142 data and age that we reported for rocks from the Nuvvuagittuq greenstone belt in Quebec, Canada. We agree that the issue of accurate mass fractionation correction is important, but stand by our discussion of this issue in our original report and our conclusion that the variation in $^{142}$Nd/$^{144}$Nd ratios reflects the decay of $^{146}$Sm caused by Sm-Nd fractionation within 300 million years of Earth's formation.

Full text at www.sciencemag.org/cgi/content/full/325/5938/267-b

### CORRECTIONS AND CLARIFICATIONS

**Reports:** "Halofuginone inhibits T$_H$17 cell differentiation by activating the amino acid starvation response" by M. S. Sundrud *et al.* (5 June, p. 1334). Ralph Mazitschek's affiliations were listed incorrectly. His correct affiliations are as follows: Center for Systems Biology, Massachusetts General Hospital, Harvard Medical School, 185 Cambridge Street, Boston, MA 02114, USA; Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02142, USA; and Chemical Biology Program, Broad Institute, Cambridge, MA 02142, USA. Also, Malcolm Whitman's e-mail address was listed incorrectly. The correct address is mwhitman@hms.harvard.edu.

**This Week in *Science*:** "Editing the genome" (15 May, p. 851). The first two sentences should have read, "The ciliate *Oxytricha trifallax* has an unusual genome with the coding regions of genes scattered through the genome. These regions are then somehow knitted together prior to their transcription and translation into proteins."

**News Focus:** "Carbon sheets an atom thick give rise to graphene dreams," by R. F. Service (15 May, p. 875). The story stated that there is no easy way to peel graphene layers off of silicon carbide wafers atop which they are sometimes grown. It should have noted that for conventional electronic applications there is no need to transfer the material, as silicon carbide can be patterned using standard semiconductor lithographic techniques.

**Letters:** "Iraq study failed replication test" by M. Spagat (1 May, p. 590). The page number in Ref. 1 should have been 1421. The page number in Ref. 2 should have been 273. The page number in Ref. 3 should have been 484.

**Letters:** "Iraq study response lacks objectivity" by F. Checchi (1 May, p. 590). The page number in Ref. 1 should have been 1421.

**Reports:** "Function of mitochondrial Stat3 in cellular respiration" by J. Wegrzyn *et al.* (6 February, p. 793). There was an inadvertent omission of lines in Figs. 1B and 2 indicating where images of the gels were spliced because samples were not placed in adjacent lanes. In Fig. 1B, the original published version failed to include the sample showing the presence of cytochrome c in heart cytoplasm; the corrected version shows this sample in the lower left-hand corner. In Fig. 2, the sample of liver mitochondrial extract and the IP of GRIM19 from this sample using antibodies specific for complex I were inadvertently switched; the corrected version shows GRIM19 with a denser band in the mitochondrial extract as compared to the IP.



Fig. 2
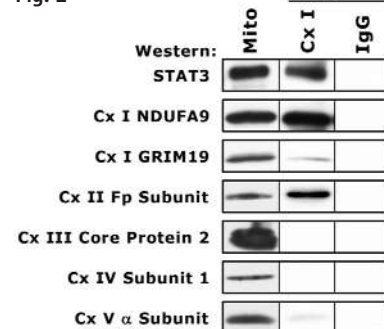


Fig. 1B
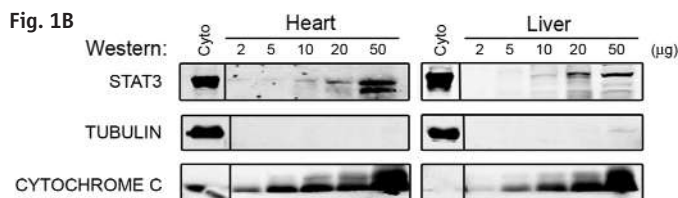
DEVELOPMENT

# The Flexible Organism

**David W. Pfennig and Cris Ledón-Rettig**

In the last chapter of *How and Why Species Multiply* (*1*), Peter and Rosemary Grant concluded that "[n]othing in evolutionary biology makes sense except in the light of ecology." In *Ecological Developmental Biology*, Scott Gilbert and David Epel argue that nothing in developmental biology makes sense except in the light of ecology either.

Traditionally, ecology—the study of the relations of organisms to one another and to their physical surroundings—has not featured prominently in developmental biology. Yet, as Gilbert and Epel (developmental biologists at Swarthmore College and Stanford University, respectively) observe, biologists have known for at least a century that ecology is an essential partner in development: In 1909, Danish biologist Wilhelm Johannsen asserted that an organism's appearance, physiology, and behavior (that is, its phenotype) derive from an interaction between its genes and its environment. Moreover, biologists long ago discovered that numerous external environmental factors—such as temperature, diet, physical stress, and the presence of predators or competitors—can alter an organism's development, often by generating a phenotype that is well suited for its current environment. For example, some plants produce large, thin leaves (which enhance photosynthetic photon harvest) in low light, and narrow, thicker leaves (which conserve water) in high light; certain insects develop wings only if they live in crowded conditions (and hence are likely to run out of adequate food in their current location). Such environmentally contingent development is so commonplace that it can be regarded as a universal property of living things.

The emerging field of ecological developmental biology (sometimes dubbed "eco-devo") explores how organisms develop and function in "real-world" environments (*2*). Analyzing development among diverse organisms under different environments is a departure from how development has previously been studied. Traditionally, research

The reviewers are at the Department of Biology, University of North Carolina, Chapel Hill, NC 27599, USA. E-mail: dpfennig@unc.edu (D.W.P.); ledonret@email.unc.edu (C.L.-R.)

has focused on a few species ("model organisms") in the laboratory (*3*). Because development was typically studied in uniform environments, past research fostered the erroneous view that environmentally contingent development is rare or unimportant. However, knowledge of ecology's role in development is essential for a complete understanding of how organisms develop and evolve. Indeed, according to Gilbert and Epel, "in addition to helping decide the survival of the fittest, the environment is also important in formulating the arrival of the fittest."

> **Ecological Developmental Biology**
> Integrating Epigenetics, Medicine, and Evolution
> *by Scott F. Gilbert and David Epel*
> Sinauer Associates, Sunderland, MA, 2009. 496 pp. $49.95. ISBN 9780878932993.

Unraveling ecology's role in development is not merely an academic exercise; it is also vital for matters of public health. Researchers have long known that certain environmental agents (including some commonly used household and agricultural products) can induce phenotypic variation by altering gene expression rather than gene nucleotide sequences. These "epigenetic" changes can cause diseases such as cancers and diabetes. In addition, because these environmental modifications can be passed stably from one generation to the next, conditions experienced by past generations can profoundly influence the health of subsequent generations. Yet, the descendants experiencing such health problems may live in a perfectly benign environment and have no (apparent)

genetic predispositions for the disease. Thus, by understanding the influences on development of both present and past environments we can gain crucial insights into the causes of disease that we might otherwise miss.

Understanding the connections between development and ecology is crucial for preserving biodiversity. In a world of increasing biological invasions, anthropogenic chemical use, and climate change, many organisms are experiencing unprecedented alterations to their environment. Such changes can cause unanticipated modifications to development, which can in turn severely affect the ecology, and even the survival, of natural populations. For instance, the pesticide DDT induces thin eggshells in birds, and atrazine (the second-largest-selling weed killer in the world) can cause sex change in many species of vertebrates (*4*). Ecological developmental biology is therefore highly relevant to conservation biology.

These are propitious and exciting times for integrating the fields of development, ecology, and evolution. Students and researchers are fortunate that (in addition to the present volume) several important books have appeared recently, including works by Carl Schlichting and Massimo Pigliucci (*5*), Mary Jane West-Eberhard (*6*), and Eva Jablonka and Marion Lamb (*7*). Nevertheless, we have only begun to construct an integrated framework. Gilbert and Epel acknowledge the arduous task ahead and "hope that college students, still relatively undifferentiated, will come up with their own connections and syntheses and that they will see patterns that we haven't yet imagined." *Ecological Developmental Biology* will serve as an excellent guide for those interested in embarking on such a synthesis. More generally, this lucid and thought-provoking book should appeal to anyone interested in under-



**Seasonal morphs.** Spring *Nemoria arizonaria* caterpillars resemble oak catkins (**left**); summer ones, newly formed twigs (**right**).

CREDIT: ERICK GREENE/UNIVERSITY OF MONTANA

standing how organisms are built, function, and evolve or how anthropogenic environmental change affects the health of ourselves and other organisms.

### References

1. P. R. Grant, B. R. Grant, *How and Why Species Multiply: The Radiation of Darwin's Finches* (Princeton Univ. Press, Princeton, NJ, 2008); reviewed in (*8*).
2. S. E. Sultan, *Trends Ecol. Evol.* **22**, 575 (2007).
3. B. Maher, *Nature* **458**, 695 (2009).
4. T. Hayes *et al.*, *Environ. Health Perspect.* **111**, 568 (2003).
5. C. D. Schlichting, M. Pigliucci, *Phenotypic Evolution: A Reaction Norm Perspective* (Sinauer, Sunderland, MA, 1998); reviewed in (*9*).
6. M. J. West-Eberhard, *Developmental Plasticity and Evolution* (Oxford Univ. Press, Oxford, 2003); reviewed in (*10*).
7. E. Jablonka, M. J. Lamb, *Evolution in Four Dimensions* (MIT Press, Cambridge, MA, 2005).
8. H. Kokko, *Science* **319**, 1187 (2008).
9. H. A. Orr, *Science* **285**, 343 (1999).
10. J. S. Levinton, *Science* **301**, 767 (2003).

PHYSICS

# Entanglement with a Twist

**Jonathan P. Dowling**

When I picked up *The Age of Entanglement*, the first thing to catch my eye was a quote on the back dustjacket, "for a moment I almost thought I understood quantum mechanics." I thought, "Oh boy, this could be trouble." Recollections danced in my noggin of uncomfortable conversations on crowded airplanes that begin with "Oh, you are a quantum physicist?— Then you must have seen *What the Bleep Do We Know*!" But proceeding through the book, my fear was never realized. I instead found a witty, charming, and accurate account of the history of that bugaboo of physics—quantum entanglement.

When I was a graduate student in physics, I made the decision to spurn a respectable career in high-energy physics theory (if nowadays one can consider superstring theory to be respectable) and embraced a future in the foundations of quantum mechanics. As Louisa Gilder repeatedly points out, in the mid-1980s such a career move was considered the kiss of death. At that time a respected professor pointedly told

The reviewer is at the Department of Physics and Astronomy, Louisiana State University, Baton Rouge, LA 70803–4001, USA. E-mail: jdowling@lsu.edu

me, "This foundations of quantum mechanics is crackpot stuff—you will never get at job." My, how times have changed.

There are many books out there on the history or foundations of quantum mechanics. Some are more technical, others more historical, but none take the unique approach that Gilder has—to focus on the quantum weirdness of entanglement itself as her book's unifying theme and to present it in an inviting and accessible way. *The Age of Entanglement* offers neither a technical nor a biographical account. Rather, as Gilder states up front, it provides a collection of reconstructed conversations among some of the 20th century's greatest physicists. These conversations all revolve around the notion of quantum entanglement: the spooky, action-at-a-distance effect predicted by quantum theory but only slowly recognized as the theory's defining feature and even more slowly shown to be experimentally verifiable.

Your opinion of the book will largely hinge on how you react to these reconstructed conversations. Concerning one such imagined conversation between Albert Einstein and Niels Bohr on a streetcar in Copenhagen, Gilder notes, "We know that the conversation … happened, because Bohr mentioned it in an interview …. The content of the conversation is easy to gather from a look at what the three men were working on … around the same time." Rather than provide dry quotations from original sources, Gilder decided to weave information from these sources into a series of imagined conversations. The author offers extensive documentation for these conversations in the notes, so they are not flights of fancy. Her technique leads to text such as, " 'If, however'—and here [Einstein] looked straight at Heisenberg, who was leaning forward in his chair, his pale hair shining in the dim room— 'as is obviously the case in modern atomic physics ….' " I suppose neither Einstein, Werner Heisenberg, nor anyone else recorded that Heisenberg's pale hair was shining in the dim room, but it makes for a good story. For this protocol to work for me, I had to first execute Coleridge's "willing suspension of disbelief" and then engage Tolkien's "secondary belief." That done, I was



**John S. Bell.** Foundational theorizer.

**The Age of Entanglement**
When Quantum Physics Was Reborn

*by Louisa Gilder*

Alfred A. Knopf, New York, 2008. 462 pp. $27.50, C$32. ISBN 9781400044177.

enthralled and found the book delightful.

Gilder skillfully relates the early discomfort physicists felt concerning some of the arcane predictions of quantum mechanics; how Einstein, Erwin Schrödinger, and others repeatedly distilled and titrated their misgiving until they were able in the 1930s to present the essence of their fears in the form of the Einstein-Rosen-Podolsky paradox; Schrödinger's cat; and the now famous notion of quantum entanglement—spooky action-at-a-distance that quantum theory must contain. Much in these older "discussions" was familiar to me from other sources. What I found most gratifying were the studiously documented dialogs of of later developments: Bell's inequalities and the consequent experiments, which proved that nature is stranger than we can think. The details of the story of David Bohm and his trials, after constructing a nonlocal hidden variable theory, were new to me. The account of John Clauser and his cohorts in the race to demonstrate (by testing Bell's theory) once and for all that this quantum weirdness did or did not exist was side-splitting. An old friend and collaborator, Clauser does in fact curse like a sailor, as Gilder often has him do. (He is a sailor, and I wonder whether the cursing or the sailing came first.) I was spellbound by the details of the struggles of Clauser and colleagues with the massive, punch-tape spewing, "coffin" contraption clanking away, day after day, in the bowels of Berkeley. It is tragic that this apparatus now lies in mothballs in the attic of LeConte Hall instead of on display at the Smithsonian.

Gilder wraps up *The Age of Entanglement* with conversations among younger quantum technologists such as Artur Ekert, Nicolas Gisin, Daniel Greenberger, Michael Horne, Terrence Rudolph, and Anton Zeilinger. As I read these pages, I wondered if I should feel slighted—there is no mention of me. Then I happened upon this description of a colleague and friend: "Meanwhile in the Sangre de Cristo Mountains of New Mexico, Paul Kwiat, an endearingly birdlike man in glasses and suspenders with boundless energy and encyclopedic knowledge, led his team in attempting various eavesdropping strategies on their Alice and Bob." Thank goodness for small favors, I thought, smiling to myself.

## ENERGY

# Beneficial Biofuels—The Food, Energy, and Environment Trilemma

Exploiting multiple feedstocks, under new policies and accounting rules, to balance biofuel production, food security, and greenhouse-gas reduction.

**David Tilman,[1]\* Robert Socolow,[2] Jonathan A. Foley,[3] Jason Hill,[3] Eric Larson,[4] Lee Lynd,[5] Stephen Pacala,[6] John Reilly,[7] Tim Searchinger,[8] Chris Somerville,[9] Robert Williams[4]**

Recent analyses of the energy and greenhouse-gas performance of alternative biofuels have ignited a controversy that may be best resolved by applying two simple principles. In a world seeking solutions to its energy, environmental, and food challenges, society cannot afford to miss out on the global greenhouse-gas emission reductions and the local environmental and societal benefits when biofuels are done right. However, society also cannot accept the undesirable impacts of biofuels done wrong.

Biofuels done right can be produced in substantial quantities (1). However, they must be derived from feedstocks produced with much

[1]Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN 55108, USA. [2]Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ 08544, USA. [3]Institute on the Environment, University of Minnesota, St. Paul, MN 55108, USA. [4]Princeton Environmental Institute, Princeton University, Princeton, NJ 08544, USA. [5]Thayer School of Engineering, Dartmouth College, Hanover, NH 03755, USA. [6]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA. [7]Center for Energy and Environmental Policy Research, MIT, Cambridge, MA 02142, USA. [8]Woodrow Wilson School, Princeton University, Princeton, NJ 08544, USA. [9]Energy Biosciences Institute, University of California Berkeley, Berkeley, CA 94720, USA.

\*To whom correspondence should be addressed: tilman@umn.edu

lower life-cycle greenhouse-gas emissions than traditional fossil fuels and with little or no competition with food production (see figure, below). Feedstocks in this category include, but may not be limited to, the following:

1) *Perennial plants grown on degraded lands abandoned from agricultural use.* Use of such lands minimizes competition with food crops. This also minimizes the potential for direct and indirect land-clearing associated with biofuel expansion, as well as the resultant creation of long-term carbon debt and biodiversity loss. Moreover, if managed properly, use of degraded lands for biofuels could increase wildlife habitat, improve water quality, and increase carbon sequestration in soils (1–3). The key to carbon gains is to use land that initially is not storing large quantities of carbon in soils or vegetation and yet is capable of producing an abundant biomass crop (4, 5). Some initial analyses on the global potential of degraded lands suggest that they could meet meaningful amounts of current global demand for liquid transportation fuels (5–7).

2) *Crop residues.* Crop residues such as corn stover and straw from rice and wheat are produced in abundance. They are rich in elements (C, N, and P) essential for maintaining soil fertility and carbon stores, and they help minimize soil erosion. Recent research suggests that it is to the benefit of farmers to leave substantial quantities of crop residues on the land (8), but that, nonetheless, even conservative removal rates can provide a sustainable biomass resource about as large as that from dedicated perennial crops grown on degraded lands (1).

3) *Sustainably harvested wood and forest residues.* Another abundant feedstock is residues from forestry operations, which include slash (branches, but not leaves or needles) that currently is left in place, unused residues from mill and pulp operations, and forest "thinnings" removed to reduce fire risk or to allow select trees to attain merchantable sizes more quickly (9, 10).

4) *Double crops and mixed cropping systems.* Double crops grown between the summer growing seasons of conventional row crops and harvested for biofuel production before row crops are planted in the spring are representative of a class of land-use options with potential to produce biofuel feedstocks without decreasing food production and without clearing wild lands (11). Mixed cropping systems in which food and energy crops are grown simultaneously present similar opportunities (12, 13).



**The best biofuels.** The search for beneficial biofuels should focus on sustainable biomass feedstocks that neither compete with food crops nor directly or indirectly cause land-clearing and that offer advantages in reducing greenhouse-gas emissions. Perennials grown on degraded formerly agricultural land, municipal and industrial sold waste, crop and forestry residues, and double or mixed crops offer great potential. The best biofuels make good substitutes for fossil energy. A recent analysis suggests that more than 500 million tons of such feedstocks could be produced annually in the United States (1).

CREDIT: M. TWOMBLY/SCIENCE

5) *Municipal and industrial wastes.* Solid waste streams, which are frequently rich in organic matter, including paper, cardboard, yard wastes, and plastics, can be converted to liquid fuels (*14*, *15*).

As global population and standards of living increase during the coming decades, both the urgency to lower greenhouse-gas emissions and the demand for transportation and meat may increase. Nonetheless, the five biomass sources discussed above—in combination with large reductions in fuel demand, achieved through increased efficiency, and large increases in both food and biomass productivity on existing farmland—could produce enough biofuels to meet a substantial portion of future energy demand for transportation (*1*).

However, looming over the future of biofuels are several wrong options. Sometimes, the most profitable way to get land for biofuels is to clear the land of its native ecosystem, be it rainforest, savanna, or grassland. The resulting release of carbon dioxide from burning or decomposing biomass and oxidizing humus can negate any greenhouse-gas benefits of biofuels for decades to centuries (*16–20*). Decisions regarding land for biofuels can have adverse consequences far beyond the land directly in question. For example, if fertile land now used for food crops (such as corn, soybeans, palm nuts, or rapeseed) is used to produce bioenergy, this could lead, elsewhere in the world, to farmers clearing wild lands to meet displaced demand for crops. In this way, indirect land-use effects of biofuels can lead to extra greenhouse-gas emissions, biodiversity loss, and higher food prices (*21*, *22*).

Dramatic improvements in policy and technology are needed to reconfigure agriculture and land use to gracefully meet global demand for both food and biofuel feedstocks. Good public policy will ensure that biofuel production optimizes a bundle of benefits, including real energy gains, greenhouse-gas reductions, preservation of biodiversity, and maintenance of food security. Present legislation in the United States takes partial steps in the right direction by specifying minimally acceptable greenhouse benefits for certain types of biofuels. Notably, the U.S. 2007 Energy Independence and Security Act states that cellulosic biofuels (such as ethanol made from cellulose) must, when both direct and indirect emission are taken into account, offer at least a 60% life-

*Dramatic improvements in policy and technology are needed to ... meet global demand for both food and biofuel feedstocks.*

cycle greenhouse-gas reduction relative to conventional gasoline (*23*).

The biofuels industry is positioned to undergo rapid growth. The attendant policy should anticipate and provide for a biofuels industry that meaningfully and positively addresses pressing sustainability and security challenges. Biofuels should receive policy support as substitutes for fossil energy only when they make a positive impact on four important objectives: energy security, greenhouse-gas emissions, biodiversity, and the sustainability of the food supply. Performance-based policies are needed that provide incentives proportional to the benefits delivered. Legislation that is vague could allow significant portions of the biofuels industry to develop along counterproductive pathways. Complementary policies must directly target related goals, such as land- and water-efficient food production, reduced agricultural greenhouse-gas emissions, and the prevention of habitat loss from land-clearing (*24*, *25*).

The recent biofuels policy dialogue in the United States is troubling. It has become increasingly polarized, and political influence seems to be trumping science. The best available science, continually updated, should be used to evaluate the extent to which various biofuels achieve their multiple objectives, and policy should reward achievement. Three steps should be taken: meaningful science-based environmental safeguards should be adopted, a robust biofuels industry should be enabled, and those who have invested in first-generation biofuels should have a viable path forward.

In support of such policy, rigorous accounting rules will need to be developed that measure the impacts of biofuels on the efficiency of the global food system, greenhouse-gas emissions, soil fertility, water and air quality, and biodiversity (*26*). Accounting rules should consider the full life cycle of biofuels production, transformation, and combustion.

Unless new technologies and life-styles are adopted globally over the coming decades, the massive projected increases in global energy and food consumption will greatly elevate atmospheric greenhouse-gas levels from fossil fuel combustion, land-clearing, and livestock production and will create immense biodiversity loss from habitat destruction and climate change. The quality of human life will be compromised. A central issue for the coming decades, then, is how the environmental impacts and potential benefits associated with meeting the global demand for food and energy can be internalized into our economic systems (*27*). This is a complex question that cannot be addressed with simplistic solutions and sound bites. It needs a new collaboration between environmentalists, economists, technologists, the agricultural community, engaged citizens, and governments around the world.

**References and Notes**
1. National Academy of Sciences, National Academy of Engineering, National Research Council, *Liquid Transportation Fuels from Coal and Biomass: Technological Status, Costs, and Environmental Impacts* (National Academy of Sciences, Washington, DC, 2009).
2. J. M. F. Johnson, A. J. Franzluebbers, S. L. Weyers, D. C. Reicosky, *Environ. Pollut.* **150**, 107 (2007).
3. G. P. Robertson *et al.*, *Science* **322**, 49 (2008).
4. K. Anderson-Teixeira, S. Davis, M. Masters, E. Delucia, *GCB Bioenergy* **1**, 75 (2009).
5. D. Tilman, J. Hill, C. Lehman, *Science* **314**, 1598 (2006).
6. C. B. Field, J. M. F. Campbell, D. B. Lobell, *Trends Ecol. Evol.* **23**, 65 (2008).
7. J. M. F. Campbell, D. B. Lobell, R. C. Genova, C. B. Field, *Environ. Sci. Technol.* **42**, 5791 (2008).
8. W. W. Wilhelm, J. M. F. Johnson, D. L. Karlen, D. T. Lightle, *Agron. J.* **99**, 1665 (2007).
9. E. D. Reinhardt, R. E. Keane, D. E. Calkin, J. D. Cohen, *For. Ecol. Manage.* **256**, 1997 (2008).
10. B. Solomon, V. Luzadis, Eds., *Renewable Energy from Forest Resources in the United States* (Routledge, New York, 2009).
11. A. H. Heggenstaller, R. P. Anex, M. Liebman, D. N. Sundberg, L. R. Gibson, *Agron. J.* **100**, 1740 (2008).
12. B. Dale, M. Allen, M. Laser, L. Lynd, *Biofuel Bioprod. Bior.* **3**, 219 (2009).
13. E. Malézieux *et al.*, *Agron. Sustain. Dev.* **29**, 43 (2009).
14. B. Antizar-Ladislao, J. L. Turrion-Gomez, *Biofuel Bioprod. Bior.* **2**, 455 (2008).
15. K. B. Cantrell, T. Ducey, K. S. Ro, P. G. Hunt, *Bioresour. Technol.* **99**, 7941 (2008).
16. F. Danielsen *et al.*, *Conserv. Biol.* **23**, 348 (2009).
17. J. Fargione, J. Hill, D. Tilman, S. Polasky, P. Hawthorne, *Science* **319**, 1235 (2008).
18. H. K. Gibbs *et al.*, *Environ. Res. Lett.* **3**, (2008).
19. M. O'Hare *et al.*, *Environ. Res. Lett.* **4**, (2009).
20. G. Piñeiro, E. G. Jobbágy, J. Baker, B. C. Murray, R. B. Jackson, *Ecol. Appl.* **19**, 277 (2009).
21. T. Searchinger *et al.*, *Science* **319**, 1238 (2008).
22. D. A. Landis, M. M. Gardiner, W. van der Werf, S. M. Swinton, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20552 (2008).
23. Energy Independence and Security Act of 2007, Public Law 110-140, H.R. 6, 2007.
24. R. Dominguez-Faus, S. E. Powers, J. G. Burken, P. J. Alvarez, *Environ. Sci. Technol.* **43**, 3005 (2009).
25. M. Wise *et al.*, *Science* **324**, 1183 (2009).
26. L. Firbank, *Bioenerg. Res.* **1**, 12 (2008).
27. J. Hill *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 2077 (2009).
28. Individuals whose backgrounds span a broad range of perspectives gathered in Princeton, NJ, to exchange views about the sustainability of biofuels, food, and the environment. After considerable back-and-forth, we arrived at the consensus presented above. We are hopeful that colleagues charged with developing biofuels policies, who are likely to span a similarly broad range of views, will benefit from our deliberations. We thank the Carbon Mitigation Initiative at the Princeton Environmental Institute, supported by BP and Ford, for funding the workshop.

MOLECULAR BIOLOGY

# Neutralizing Toxic RNA

Oligonucleotides that target aggregates of RNA and protein show potential for treating myotonic dystrophy.

**Thomas A. Cooper**

Expansion of tandem nucleotide repeats in the human genome—most often the triplets CAG and CTG—causes a number of inherited diseases such as Huntington's disease, spinocerebellar ataxia, and myotonic dystrophy. An expansion within the coding region of a gene can reduce the normal function of the corresponding protein and/or cause an aberrant function (*1*), whereas expansions in noncoding regions can cause disease (such as myotonic dystrophy) without interfering with protein sequence. In myotonic dystrophy, RNA that is transcribed from DNA containing noncoding expansions is causative of disease pathogenesis. On page 336 of this issue, Wheeler *et al.* (*2*) present a potentially highly effective approach to neutralize the toxic effects of these RNAs.

The predominant form of myotonic dystrophy in the United States, DM1, is caused by expansion of a CTG repeat located in the 3′untranslated region (3′UTR) of the *DMPK* gene, from a normal range of 5 to 35 repeats to a pathogenic range of 80 to >2500 repeats (*3*). The expanded allele is transcribed to produce RNA containing expanded CUG repeats (CUG$^{exp}$) that becomes stuck in nuclear foci, precluding its export to the cytoplasm for translation into DMPK protein (*4*). Although loss of DMPK contributes to the disease, toxicity of CUG$^{exp}$ RNA plays the major role. CUG$^{exp}$ RNA folds into an imperfect hairpin structure (*5*) that resembles the natural binding site for the protein muscleblind-like 1 (MBNL1). MBNL is consequently sequestered by the RNA, not only resulting in loss of its normal function in RNA splicing, but enhancing formation of foci that trap CUG$^{exp}$ RNA in the nucleus (*6*, *7*). Another component of pathogenicity is aberrant activation of protein kinase C, which leads to increased activity of a second splicing regulator, CUG-binding protein 1 (CUGBP1) (*8*). Both MBNL1 and CUGBP1 coordinately regulate the alternative splicing of precursor mRNA during development. CUG$^{exp}$ RNA
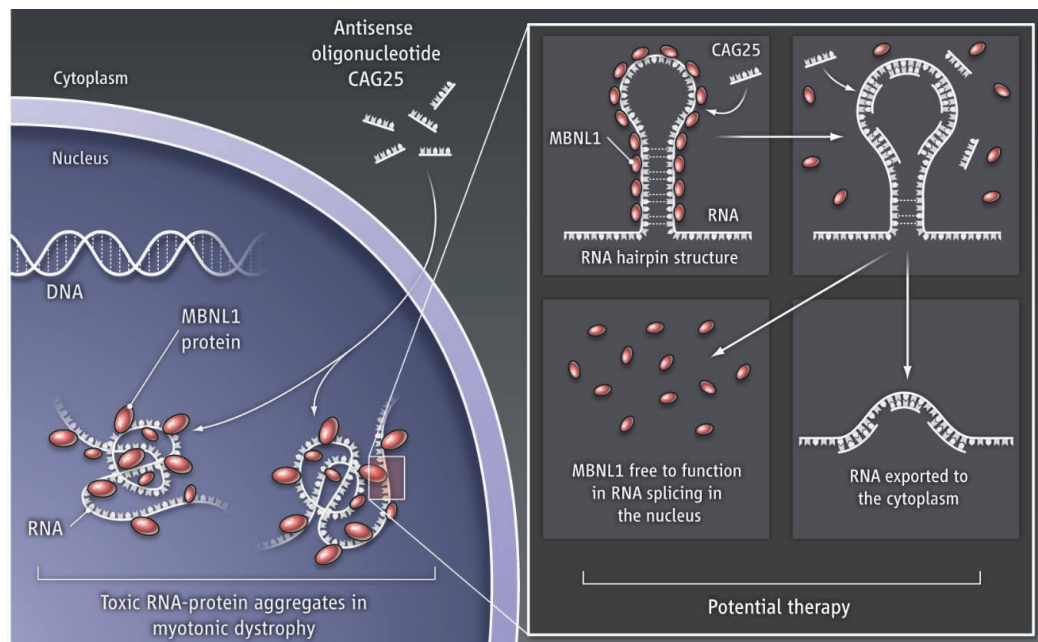


**Handling repeats.** An oligonucleotide (CAG25) frees an RNA binding protein (MBNL1) from toxic nuclear aggregates. The aggregates contain RNA with tandem trinucleotide repeats. In myotonic dystrophy, a gene encoding the protein DMPK can contain up to 2500 CUG repeats. Its corresponding RNA is implicated in disease pathogenesis.

disrupts this program, resulting in the aberrant expression of embryonic splicing patterns in adult tissues. One of the best-characterized misregulated splicing events in DM1 is of the RNA encoding the muscle-specific chloride channel (CLCN1). This results in skeletal muscle hyperexcitability, causing the myotonia for which the disease is named (*9*).

The solution proposed by Wheeler *et al.* is to use a modified (morpholino) DNA antisense oligonucleotide (CAG25, which is 25 nucleotides in length) to form base pairs with the CUG$^{exp}$ RNA, promote unwinding of the hairpin, and block access to RNA binding proteins (see the figure). The proof of principle was to test the effects of the oligonucleotide in a mouse model of DM1, in which expression of a transgene, containing 250 tandem CTGs in the 3′UTR, reproduces DM1-like features in skeletal muscle (*10*). The approach of "neutralizing" the corresponding RNA is a winning strategy in three ways. MBNL1 is released from the RNA, reversing several splicing abnormalities; for instance, reversal of the CLCN1 RNA splicing defect resulted in the loss of myotonia in the mouse model. CAG25 also caused nuclear foci to dissipate, thus increasing the amount (and translation) of CUG$^{exp}$ RNA in the cytoplasm.

This suggests the potential to reverse features that could result from *DMPK* haploinsufficiency in the human disease. Also, the overall amount of CUG$^{exp}$ RNA was reduced by 50%. Although this still leaves a pathogenic amount of RNA in the cell, it points to the potential for further manipulation. The mechanism that reduces the amount of CUG$^{exp}$ RNA remains unclear, but breaking up the RNA-protein complex may simply allow RNA degradation by endogenous processes.

Indeed, modified oligonucleotides are in clinical trials to treat a variety of diseases (*11*). As with all therapeutics, delivery is an issue, particularly for multisystemic diseases such as DM1, which severely affects skeletal muscle, heart, and brain. Wheeler *et al.* used injection combined with electroporation in skeletal muscle to achieve long-lasting effects of CAG25 in mice (a single injection reversed splicing abnormalities for at least 14 weeks). However, studies in Duchenne muscular dystrophy demonstrated the importance of using systemic delivery to treat multiple tissues simultaneously, because improved skeletal muscle function can further challenge heart function, which is also compromised by the disease (*12*).

CUG$^{exp}$ RNA also activates protein kinase C, and it will be of interest to deter-

Departments of Pathology and Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030, USA. E-mail: tcooper@bcm.edu

CREDIT: C. BICKEL/SCIENCE

mine whether the oligonucleotide will ame-liorate this effect. Another issue is whether redistribution of CUG$^{exp}$ RNA introduces a different toxicity; RNA binding proteins that function in the nucleus may also have cyto-plasmic RNA processing functions (*13*). This is the case for CUGBP1 and at least one member of the MBNL family (*14*, *15*). Because the mutant gene will produce toxic CUG$^{exp}$ RNA for the life of the individual, treatment with oligonucleotides could last for decades and it might be best to get rid of the CUG$^{exp}$ RNA altogether. One approach would be combined therapies such as the CAG25 oligonucleotide to release the RNA from nuclear foci and another differently

modified DNA oligonucleotide that allows RNA degradation by endogenous ribonu-clease H (*16*). This double-pronged attack could produce additive if not synergistic molecular effects and a desirable long-term therapeutic result.

### References

1. H. T. Orr, H. Y. Zoghbi, *Annu. Rev. Neurosci.* **30**, 575 (2007).
2. T. M. Wheeler *et al.*, *Science* **325**, 336 (2009).
3. P. S. Harper, *Myotonic Dystrophy* (Saunders, London, ed. 3, 2001).
4. B. M. Davis, M. E. McCurrach, K. L. Taneja, R. H. Singer, D. E. Housman, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7388 (1997).
5. M. Napierala, W. J. Krzyosiak, *J. Biol. Chem.* **272**, 31079 (1997).
6. W. Dansithong, S. Paul, L. Comai, S. Reddy, *J. Biol. Chem.* **280**, 5773 (2005).
7. J. W. Miller *et al.*, *EMBO J.* **19**, 4439 (2000).
8. N. Kuyumcu-Martinez, G. S. Wang, T. A. Cooper, *Mol. Cell* **28**, 68 (2007).
9. T. M. Wheeler, J. D. Lueck, M. S. Swanson, R. T. Dirksen, C. A. Thornton, *J. Clin. Invest.* **117**, 3952 (2007).
10. A. Mankodi *et al.*, *Science* **289**, 1769 (2000).
11. L. Bonetta, *Cell* **136**, 581 (2009).
12. D. Townsend, S. Yasuda, S. Li, J. S. Chamberlain, J. M. Metzger, *Mol. Ther.* **16**, 832 (2008).
13. A. B. Shyu, M. F. Wilkinson, *Cell* **102**, 135 (2000).
14. I. A. Vlasova *et al.*, *Mol. Cell* **29**, 263 (2008).
15. Y. Adereth, V. Dammai, N. Kose, R. Li, T. Hsu, *Nat. Cell Biol.* **7**, 1240 (2005).
16. S. I. Rudnick, J. Swaminathan, M. Sumaroka, S. Lieb-haber, A. M. Gewirtz, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13787 (2008).

---

## CLIMATE CHANGE

# What Drives Climate Flip-Flops?

**Axel Timmermann and Laurie Menviel**

*A numerical modeling study questions the validity of a key paradigm in rapid climate change studies.*

The threat of rapid climate change con-cerns politicians and climate scien-tists alike. To assess whether abrupt transitions in the climate system are likely to occur in coming decades and centuries, an in-depth understanding of the conditions that led to past abrupt climate changes is required. On page 310 of this issue, Liu *et al.* (*1*) use a state-of-the-art general circula-tion model to investigate a poster child for dramatic climate events in recent Earth his-tory: the Bølling-Allerød transition (see the figure). The results suggest that this transi-tion can be simulated simply as the North Atlantic climate response to a rapidly ceas-ing glacial meltwater flow.

Around 14,600 years ago, the atmospheric circulation over the North Atlantic region flipped within just a few years to another state (*2*); also, Greenland temperatures skyrock-eted by >10°C over several decades (*3*), termi-nating a cold phase known as Heinrich Event 1. The global impacts of this Bølling-Allerød transition have been well documented with climate proxy records such as sediment cores and ice cores, but the physical conditions that triggered the transition remain controversial. The temperature evolution from the Heinrich Event 1 to the Bølling-Allerød and the sub-sequent Younger Dryas cold phase (see the figure) is strikingly similar to the Dansgaard-Oeschger cycles (*4*) that dominated North-ern Hemispheric climate between 60,000 and

30,000 years ago (*5*). Hence, unraveling the processes that triggered the Bølling-Allerød transition may also help to elucidate the mys-terious, tantalizingly regular Dansgaard-Oeschger cycles (*6*).

In one of the most epic numerical mod-eling efforts of the climate community to date, Liu *et al.* have now set out to simulate the period from the Last Glacial Maximum (21,000 years ago) through the Heinrich event I and the Bølling-Allerød transition



**North Atlantic sea-surface temperature evolution.** The simulated North Atlantic temperature in the Iberian Margin region (*1*) agrees well with a reconstructed temperature index obtained by averaging different alk-enone-based sea-surface temperature reconstructions from cores SU-8118 (*13*) and MD01-2443 (*14*). The Bølling-Allerød warming can be decom-posed into contributions that originate from the recovery of the Atlantic meridional overturning circulation (AMOC), an overshooting effect, and the climate response to increasing greenhouse gas concentration and shrinking glacial ice sheets.

(see the figure). Prescribing the observed history of greenhouse gas concentrations from 22,000 years ago to 14,000 years ago, as well as time-varying solar insolation due to astronomical effects and reconstructions of ice-sheet topography and meltwater dis-charge, the equations of one of the most sophisticated climate models (CCSM3) were solved numerically on the Jaguar supercom-puter at the Oak-Ridge National Laboratory. After about one-and-a-half years of num-ber crunching, the first results (*1*) provide an unprecedented view into the mechanisms of the last glacial termination and the emergence of the Bølling-Allerød transi-tion. The findings call for a paradigm shift in our understanding of abrupt climate change.

Inspired by theoreti-cal considerations (*7*) and idealized climate modeling results (*8*), pre-vious attempts to explain abrupt millennial-scale climate change during the last glacial period—including Dansgaard-Oeschger events and the Bølling-Allerød transi-tion—have relied on the idea that for a wide range of North Atlantic fresh-water inputs, the Atlantic

International Pacific Research Center, School of Ocean and Earth Science and Technology (SOEST), University of Hawaii, Honolulu, HI 96822, USA. E-mail: axel@hawaii.edu; menviel@hawaii.edu

meridional overturning circulation (AMOC) has two different stable circulation regimes (9). According to this paradigm, abrupt climate change emerges as a flip-flopping between these equilibrium states. Liu et al., however, question the mere existence of this ocean "ambivalence" under glacial conditions.

By prescribing a Northern Hemispheric freshwater forcing scenario in which the discharge of meltwater from the retreating glacial ice sheets during Heinrich Event 1 stops suddenly, Liu et al. are able to simulate an abrupt recovery of the AMOC that triggers the transition from Heinrich Event 1 conditions to the Bølling-Allerød. The results are in good agreement with paleoclimate reconstructions based on climate proxy records. The rapid AMOC recovery described by Liu et al. also involves an overshooting effect (see the figure) that was noted in previous climate model simulations (10) and that affects mostly high-latitude climate in the Northern Hemisphere. Another contributor to the Bølling-Allerød warming is a 40 parts per million by volume increase in atmospheric carbon dioxide that accompanied Heinrich Event 1 (11) and further accelerated the deglaciation. Its origin remains a mystery.

The Liu et al. study initiates a new era of paleoclimate modeling. Previous paleoclimate modeling studies using state-of-the-art climate models adopted the time-slice concept to explore, for example, the climate of the Last Glacial Maximum or the mid-Holocene Optimum. In contrast, the transient modeling approach of Liu et al. simulates the climate evolution since the Last Glacial Maximum by prescribing the time evolution of the external boundary conditions based on astronomical theory, ice-sheet reconstructions, and the history of greenhouse gas concentrations. It thus offers the unique possibility to study the full spatiotemporal behavior of climate change, including the mechanisms of abrupt climate change, and to directly compare the resulting temporal features with paleoclimate data—for example, from sediment cores and ice cores.

However, such ambitious projects are computationally very demanding, and it might take several years before similar transient simulations can be carried out routinely with other coupled general circulation models. Even completing the CCSM3 simulation (1) by running it into the present will require another 2 to 3 million CPU hours on the Jaguar supercomputer.

Climate models are tools to promote our understanding of the climate system and pre-dict its future evolution. Efforts to elucidate the complex mechanisms of past climate change will rely not only on the most realistic models such as the CCSM, but also on Earth system models of intermediate complexity that can explore Earth's long-term climate dynamics (12), and on conceptual models that have often initiated major shifts in our understanding of climate. Ultimately, breakthroughs in our understanding of Earth's climate evolution will come from close interactions between paleoproxy experts, paleoclimate modelers, and climate dynamicists. It is time to train a new generation of scientists familiar with all these fields.

**References**
1. Z. Liu et al., Science **325**, 310 (2009).
2. J. P. Steffensen et al., Science **321**, 680 (2008).
3. J. P. Severinghaus, E. J. Brook, Science **286**, 930 (1999).
4. S. J. Johnsen et al., Nature **359**, 311 (1992).
5. M. Schulz, A. Paul, A. Timmermann, Geophys. Res. Lett. **29**, 2193 (2002).
6. M. Schulz, Paleoceanography **17**, 1014 (2002).
7. H. Stommel, Tellus **13**, 224 (1961).
8. S. Manabe, R. J. Stouffer, Clim. Dyn. **1**, 841 (1988).
9. T. F. Stocker, Quat. Sci. Rev. **19**, 301 (2000).
10. A. Ganopolski, S. Rahmstorf, Nature **409**, 153 (2001).
11. E. Monnin, Science **291**, 112 (2001).
12. O. Timm, A. Timmermann, J. Clim. **20**, 4377 (2007).
13. E. Bard et al., Science **289**, 1321 (2000).
14. B. Martrat et al., Science **317**, 502 (2007).

---

EPIDEMIOLOGY

# Does Viral Diversity Matter?

Graham F. Medley[1] and D. James Nokes[1,2]

Detailed analyses of population demographics and viral diversity should help predict the impact of childhood vaccination for endemic viruses such as rotavirus.

Epidemic viruses, such as severe acute respiratory syndrome (SARS) and influenza A, cause diseases that rapidly spread to many people, and seem to attract more scientific and public attention than do endemic viruses, which are continually present in populations. Yet endemic viruses cause far greater disability and death. But epidemic viruses are endemic somewhere, or will become so, and endemic viruses are often recurrently epidemic. So developing a full understanding of the mechanisms that promote and drive endemicity is key to reducing the overall burden of viral disease and reducing the risk of future, widespread pandemics. On page 290 of this issue, Pitzer et al. (1) investigate the causes of epidemics of rotavirus, a major, global endemic virus.

Rotavirus and respiratory syncytial virus are endemic viruses with epidemic propensity. They infect the epithelia of the gastrointestinal tract and respiratory tract, respectively, and cause severe childhood diseases. Both are RNA viruses (with a high propensity for genetic and antigenic variation) that occur in regular (mostly annual) epidemics (see the figure), infect children at a young age, and reinfect people throughout their lives. Disease is primarily associated with the first infection—acquired immunity reduces the severity of subsequent infections—but repeat infection is common and facilitated in part by viral antigenic diversity. New variants of rotavirus arise through mixing of its segmented genes during superinfection (simultaneous infection by two viral strains).

Innate and acquired immune responses severely limit survival of rotavirus and respiratory syncytial virus, so for a virus to persist, it must be transmitted to new, susceptible



Weekly number of laboratory-confirmed rotavirus reports in England and Wales, 1984-2007

**Rotavirus regularity.** Rotavirus occurs in annual epidemics. Shown are the weekly number of rotavirus reports in England and Wales from 1984 to 2007. The dashed line indicates that in 1999, most labs switched to antibody-based testing. The graph is taken from (8) with permission from Eurosurveillance.
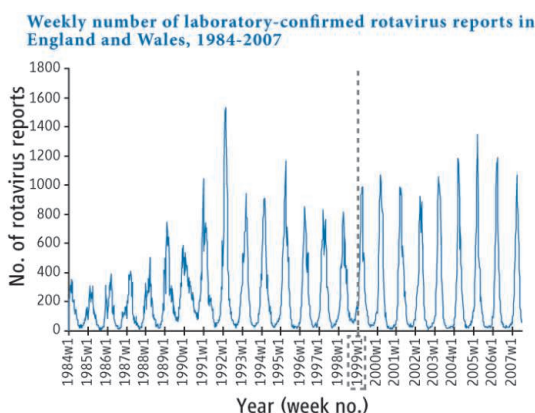
[1]Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK. [2]Kenya Medical Research Institute-Wellcome Trust Research Programme, Kilifi, Kenya. E-mail: graham.medley@warwick.ac.uk

hosts within a population. Generally, there are two sources of susceptible hosts: births and waning immunity. The latter is a manifestation of either short-lived protection or viruses with different antigenic structure so that hosts who are immune to different viral types can be reinfected. Consequently, viruses that generate long-lasting immunity and cannot generate sufficient diversity within a population are epidemic. For example, measles, mumps, and rubella have very limited antigenic diversity and must wait for births to accumulate before an epidemic can sweep through the susceptible cohort (2). But if a virus has sufficient antigenic diversity to enable it to continually reinfect, then why should it be epidemic, and not maintain a constant prevalence all year round?

Pitzer *et al.* address this apparent paradox by showing that births play a role in generating epidemics of rotavirus, and use this result to predict the impact of vaccination. A corollary of their findings is that first infections (not reinfections) are intrinsically important to viral persistence in a population, so that reducing these to negligible levels through mass childhood vaccination will eliminate an endemic virus.

However, one cannot firmly conclude that infection of susceptible hosts (newborns) is chiefly responsible for driving the intrinsic epidemic cycle of rotavirus or respiratory syncytial virus. There is a dynamic interaction between host demography and viral diversity that determines epidemiology (3),

so that epidemics may be created, or at least influenced, by strain variation (4). In developing countries, rotavirus is less seasonal than in the United States and Europe, perhaps influenced by higher birth rates. But the very high diversity of co-circulating rotavirus variants in Africa and other developing countries (5) could indicate that reduced seasonality results from more continuous reinfections by antigenically different variants. Furthermore, contact patterns (that determine which viruses are circulating in a particular subgroup) vary (6), and are likely to be different in developing countries (although specific data are currently lacking) and to vary with social circumstances and situation, including birth rate and contact between children and adults. This is a complex situation about which we understand little, although the impact of vaccination will be revealing.

There are two possible, general outcomes to vaccination (7). One is that a vaccine will effectively reduce viral prevalence, disease, and diversity. If first infections are critical for rotavirus persistence, then reinfections and viral diversity are essentially bystanders. The other outcome is that the vaccine will reduce disease, but viral prevalence will remain unchanged. If rotavirus can survive in a population of already infected hosts, then, although diversity might be altered, it will remain high, and prevalence unaltered.

Current approaches to vaccines, particularly live-attenuated vaccines, may be less effective

in malnourished populations with high rates of infection (and superinfection) and may also be compromised by the presence of maternal antibodies or immunological immaturity, so that very young children are less easy to protect. If there are substantial vaccine failures or the proportion of the population that is vaccinated is low, or if there is a need to vaccinate older age groups, then continued circulation of virus is unwelcome. So if the last situation pertains, then there is a rationale for developing sterilizing vaccines that prevent reinfections.

Endemic infections generally are well adapted to their environments, which extend from the biologic into the economic, social, and political spheres. Because RNA viruses such as rotavirus and respiratory syncytial virus are highly adaptive, changes (such as birth rates) and interventions (such as vaccination) will have long-term consequences that are difficult to predict and might be serious.

**References and Notes**
1. V. E. Pitzer *et al.*, *Science* **325**, 290 (2009).
2. B. T. Grenfell, O. N. Bjornstad, J. Kappey, *Nature* **414**, 716 (2001).
3. B. T. Grenfell *et al.*, *Science* **303**, 327 (2004).
4. L. J. White, M. Waris, P. A. Cane, D. J. Nokes, G. F. Medley, *Epidemiol. Infect.* **133**, 279 (2005).
5. N. Santos, Y. Hoshino, *Rev. Med. Virol.* **15**, 29 (2005).
6. J. Mossong *et al.*, *PLoS Med.* **5**, e74 (2008).
7. M. G. M. Gomes, L. J. White, G. F. Medley, *J. Theor. Biol.* **228**, 539 (2004).
8. C. J. Atchison *et al.*, *Eurosurveillance* **14**, 19217 (2009).
9. Financial support was provided by the Wellcome Trust (grants 061584 and 076278).

## PHYSICS

# Is Quantum Theory Exact?

Stephen L. Adler[1] and Angelo Bassi[2]

Future experiments may tell us if quantum mechanics is an approximation to a deeper-level theory.

Quantum mechanics has enjoyed many successes since its formulation in the early 20th century. It has explained the structure and interactions of atoms, nuclei, and subnuclear particles, and has given rise to revolutionary technologies, such as integrated circuit chips and magnetic resonance imaging. At the same time, it has generated puzzles that persist to this day.

These puzzles are largely connected with the role of measurements in quantum mechanics (1). According to the standard quantum postulates, given the total energy (the Hamilto-

nian) of a quantum system, the state of the system (the wave function) evolves with time in a predictable, deterministic way as described by Schrödinger's equation. However, when a physical quantity—the quantum mechanical spin, for example—is "measured," the outcome is not predictable. If the wave function contains a superposition of components, such as spin-up and spin-down (each with a definite spin value, weighted by coefficients $c_{up}$ and $c_{down}$), then each run gives a definite outcome, either spin-up or spin-down. But repeated experimental runs yield a probabilistic distribution of outcomes. The outcome probabilities are given by the absolute value squared of the corresponding coefficient in the initial wave function. This recipe is the Born rule.

How can we reconcile this probabilistic distribution of outcomes with the determinis-

tic form of Schrödinger's equation? What precisely constitutes a "measurement?" At what point do superpositions break down, and definite outcomes appear? Is there a quantitative criterion, such as size of the measuring apparatus, governing the transition from coherent superpositions to definite outcomes? These puzzles have inspired a large literature in physics and philosophy.

There are two distinct approaches. One is to assume that quantum theory is exact, but that the interpretive postulates must be modified to eliminate apparent contradictions. The second approach is to assume that quantum mechanics is not exact, but instead is a very accurate approximation to a deeper-level theory that reconciles the deterministic and probabilistic aspects. This may seem radical, even heretical, but looking back in the history
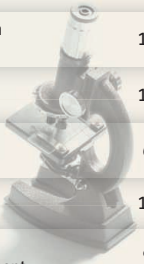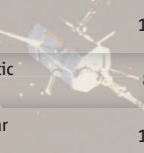
[1]Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA. E-mail: adler@ias.edu [2]Department of Theoretical Physics, University of Trieste, Strada Costiera 11, 34014 Trieste, Italy, and Istituto Nazionale di Fisica Nucleare, Trieste Section, Via Valerio 2, 34127 Trieste, Italy. E-mail: bassi@ts.infn.it

of physics, there are precedents. Newtonian mechanics was considered to be exact for several centuries, before it was supplanted by relativity and quantum theory. But apart from this history, there is another important motivation for considering modifications of quantum theory. Having an alternative theory, to which current and proposed experiments can be compared, allows a quantitative measure of the accuracy to which quantum theory can be tested.

We focus here on phenomenological approaches that modify the Schrödinger equation. A successful phenomenology must accomplish many things: It must explain why repetitions of the same measurement lead to definite, but differing, outcomes, and why the probability distribution of outcomes is given by the Born rule; it must permit quantum coherence to be maintained for atomic and mesoscopic systems, while predicting definite outcomes for measurements with realistic apparatus sizes in realistic measurement times; it should conserve overall probability, so that particles do not spontaneously disappear; and it should not allow superluminal transmission of signals.

Over the past two decades, a phenomenology has emerged that satisfies these requirements. One ingredient is the observation that rare modifications, or "hits," acting on a system by localizing its wave function, do not alter coherent superpositions for microscopic systems, but when accumulated over a macroscopic apparatus can lead to definite outcomes that differ from run to run (2). A second ingredient is the observation that the classic "gambler's ruin" problem in probability theory gives a mechanism that can explain the Born rule governing outcome probabilities (3). Suppose that Alice and Bob each have a stack of pennies, and flip a fair coin. If the coin shows heads, Alice gives Bob a penny, while if the coin shows tails, Bob gives Alice a penny. The game ends when one player has all the pennies and the other has none. Mathematical analysis shows that the probability of each player winning is proportional to the size of their initial stack of pennies. By mapping the initial stack sizes into the modulus squared of the initial spin component coefficients ($c_{up}$ and $c_{down}$), and the random flips of the fair coin into the random "hits" acting on the wave function, one then has a mechanism for obtaining the Born rule.

The combination of these two ideas leads to a definite model, called the continuous spontaneous localization (CSL) model (4), in which a Brownian motion noise term coupled nonlinearly to the local mass density is added to the Schrödinger equation. This noise is responsi-



**Upper bounds on $\lambda$**

| Laboratory experiments | Decades above the conventional value |
|---|---|
| Fullerene diffraction experiments | 13 |
| Decay of supercurrents | 14 |
| Spontaneous x-ray emission from Ge | 6 |
| Proton decay | 18 |
| Mirror cantilever interferometric experiment | 9 |

| Cosmological data | Decades above the conventional value |
|---|---|
| Dissociation of cosmic hydrogen | 17 |
| Heating of intergalactic medium (IGM) | 8 |
| Heating of interstellar dust grains | 15 |

**Quantum boundaries**. Upper bounds on $\lambda$ obtained from laboratory experiments and cosmological data, compared with the conventional CSL model value $\lambda$ $\sim 10^{-17} s^{-1}$ (with noise correlation length, $r_C = 10^{-5}$ cm). Reducing the numbers by 8 gives the distance of each bound from the enhanced value $\lambda \sim 10^{-9} s^{-1}$ obtained if one assumes that latent image formation constitutes measurement.

ble for the spontaneous collapse of the wave function. At the same time, the standard form of this model has a linear evolution equation for the noise-averaged density matrix, forbidding superluminal communication. Other versions of the model exist (5, 6), and an underlying dynamics has been proposed for which this model would be a natural phenomenology (7).

The CSL model has two intrinsic parameters. One is a rate parameter, $\lambda$, with dimensions of inverse time, governing the noise strength. The other is a length, $r_C$, which can be interpreted as the spatial correlation length of the noise-field. Conventionally, $r_C$ is taken as $10^{-5}$ cm, but any length a few orders of magnitude larger than atomic dimensions ensures that the "hits" do not disrupt the internal structure of matter. The reduction rate in the CSL model is the product of the rate parameter, times the square of the number of nucleons (protons and neutrons) within a correlation length that are displaced by more than this length, times the number of such displaced groups. Applying this formula, and demanding that a minimal apparatus composed of $\sim 10^{15}$ nucleons should settle to a definite outcome in $\sim 10^{-7}$ s or less, with the conventional $r_C$, requires that $\lambda$ should be greater than $\sim 10^{-17}$ s$^{-1}$ (4, 5). If one requires that latent image formation in photography, rather than subsequent development, consti-

tutes a measurement, the fact that only 5000 or so nucleons move appreciable distances in a few hundredths of a second in latent image formation requires an enhanced lower bound for $\lambda$ a factor of $\sim 10^8$ larger (8).

An upper bound on $\lambda$ is placed by the requirement that apparent violations of energy conservation, taking the form of spontaneous heating produced by the noise, should not exceed empirical bounds, the strongest of which comes from heating of the intergalactic medium (8). Spontaneous radiation from atoms places another stringent bound (9), which can, however, be evaded if the noise is nonwhite, with a frequency cutoff (10–12). Laboratory and cosmological bounds on $\lambda$ (for $r_C = 10^{-5}$ cm) are summarized in the figure, which gives for each bound the order of magnitude improvement needed to confront the conventional CSL model value of $\lambda$.

Accurate tests of quantum mechanics that have been performed or proposed include diffraction of large molecules in fine mesh gratings (13) and a cantilever mirror incorporated into an interferometer (14). The figure shows the current limit on $\lambda$ that has been obtained to date in fullerene diffraction and the limit that would be obtained if the proposed cantilever experiment attains full sensitivity (15). To confront the conventional (enhanced) value of $\lambda$, one would have to diffract molecules a factor of $10^6$ ($10^2$) larger than fullerenes.

Experiments do not yet tell us whether quantum theory is exact or approximate. Future lines of research include refining the sensitivity of current experiments to reach the capability of making this decision and achieving a deeper understanding of the origin of the CSL noise field.

**References**
1. A. J. Leggett, *Science* **307**, 871 (2005).
2. G. C. Ghirardi, A. Rimini, T. Weber, *Phys. Rev. D Part. Fields* **34**, 470 (1986).
3. P. Pearle, *Phys. Rev. D Part. Fields* **13**, 857 (1976).
4. G. C. Ghirardi, P. Pearle, A. Rimini, *Phys. Rev. A* **42**, 78 (1990).
5. A. Bassi, G. C. Ghirardi, *Phys. Rep.* **379**, 257 (2003).
6. P. Pearle, in *Open Systems and Measurements in Relativistic Quantum Field Theory* (Lecture Notes in Physics, vol. 526), H.-P. Breuer, F. Petruccione, Eds. (Springer, Berlin, 1999), pp. 195–234.
7. S. L. Adler, *Quantum Theory as an Emergent Phenomenon* (Cambridge Univ. Press, Cambridge, 2004).
8. S. L. Adler, *J. Phys. A* **40**, 2935 (2007).
9. Q. Fu, *Phys. Rev. A* **56**, 1806 (1997).
10. S. L. Adler, F. M. Ramazanoğlu, *J. Phys. A* **40**, 13395 (2007).
11. A. Bassi, S. L. Adler, *J. Phys. A* **40**, 15083 (2007).
12. A. Bassi, S. L. Adler, *J. Phys. A* **41**, 395308 (2008).
13. M. Arndt et al., *Nature* **401**, 680 (1999).
14. W. Marshall, C. Simon, R. Penrose, D. Bouwmeester, *Phys. Rev. Lett.* **91**, 130401 (2003).
15. A. Bassi, E. Ippoliti, S. L. Adler, *Phys. Rev. Lett.* **94**, 030401 (2005).

PHOTO CREDIT: JUPITERIMAGES

COMPUTER SCIENCE

# Toward a Smarter Web

*Interactive evolutionary algorithms are increasingly implemented in Web sites to respond to user preferences.*

Gregory S. Hornby[1] and Tolga Kurtoglu[2]

Since its creation in the early 1990s, the World Wide Web has evolved from its initial, static Web sites to the dynamic, interactive Web sites of today. But whereas existing Web sites merely respond directly to user input, there is growing interest in making them adaptive through the use of computational intelligence. A promising approach for this involves a family of optimization techniques called evolutionary algorithms.

A typical evolutionary algorithm (*1*) starts with an initial population of randomly generated, digital solutions to a problem ("individuals"). These individuals are evaluated with a user-supplied fitness function; on the basis of their fitness scores, better individuals are stochastically selected to act as "parents." Either one parent is copied while making a small change to it (mutation), or parts of two parents are combined to make a new individual (recombination). This breeding process is repeated for a fixed number of evaluations or until the problem has been solved.

Since the 1990s, evolutionary algorithms have been applied to architectural problems from arch dams and suspension bridges to building plans. In industrial and engineering design, they have found use in color design for knitwear, shape design for scissors, and car body styling, as well as for creating complex devices such as gyroscopes and wind turbines. Examples are the nose cone of Hitachi's Series N700 Bullet Train (*2*) and the communications antennas for the spacecraft in NASA's ST-5 mission, a test mission to validate new space technologies and study the magnetosphere (*3*).

Traditional evolutionary algorithms optimize against explicit fitness functions, but problems involving taste or aesthetics cannot be easily reduced to a mathematical function of goodness. Instead, a human user can perform evaluation manually, as first proposed by Dawkins (*4*). Running on a standard PC, the first interactive evolutionary algorithm showed the user computer-generated images, of which the user would select one as the parent for the next generation. By iteratively selecting images on the basis of aesthetics, the algorithm produces more and more visually appealing images over time. This has become the standard interface for interactive evolutionary algorithms (see the figure).

Since then, interactive evolutionary algorithms have been used in various human-computer interactive design systems. Most applications are visual, such as the evolution of images (*5*), three-dimensional shapes (*6*), and architectural forms (*7*), but they have also been used for musical tasks, including sound synthesis and composition. For example, GenJam is an interactive evolutionary algorithm for real-time jazz improvisation (*8*). Along with its creator, Al Biles, it forms a virtual jazz quintet that has performed at more than 100 private receptions.

The first Web browser that supported images (Mosaic) was released in April 1993. Soon afterward, the first online interactive evolutionary algorithms appeared. The International Interactive Genetic Art 1 (IIGA1) (*9*) and its successor, IIGA2, had more than 100,000 visitors who collectively created thousands of images over a period of 4 years. In an early commercial application, Affinnova



**Learning by doing it.** An explicit interactive evolutionary algorithm presents the user with several computer-evolved design options (**A**). On the basis of the user's selection (red outline), a new set of evolved designs is presented (**B**).

(www.affinnova.com) has used an interactive evolutionary algorithm–based system to design product packaging since 2000. Nymbler (www.nymbler.com) allows users to evolve baby names instead of images.

A key challenge for interactive evolutionary algorithms is user fatigue (*10*). For typical noninteractive evolutionary algorithms, tens of thousands of evaluations are needed to achieve interesting results—orders of magnitude more than can be expected from a single user. On the Web, many users are available, but even this multiplier effect may not overcome user fatigue: Because the interactions are distributed in time, no single user is likely to experience evolution at a sufficiently fast pace for it to be interesting.

[1]University of California at Santa Cruz, University Affiliated Research Center, Mail Stop 269-3, Moffett Field, CA 94035, USA. [2]Mission Critical Technologies Inc., NASA Ames Research Center, Moffett Field, CA 94035, USA. E-mail: gregory.s.hornby@nasa.gov; tolga.kurtoglu@nasa.gov

Given that user fatigue limits the number of interactions, one must make the most of what little data the user provides. The main approach to this is to automate most design evaluations and only selectively query the user. For parameterized design spaces (*11*), function approximation techniques can be used to assign a goodness score, based on similar designs evaluated by the user. But creating an adequate approximation is difficult, and this approach does not generalize to more open-ended, generative representations (*11*) for encoding designs. Another approach—using mathematical heuristics of aesthetics—has found some success in the interactive evolution of jewelry (*12*). The most promising long-term approach is to continuously learn and refine a model of user preferences (*13*) while simultaneously using this model to perform most evaluations.

The interactive systems described above explicitly present the user with choices to select from. Interactive evolutionary algorithms on the Web can also be invisible to the user. For example, the company SnapAds (www.snapads.com) uses an implicit interactive evolutionary algorithm to evolve banner ads. Variations of an ad are placed on Web pages. On the basis of click-through rates, the ad layout evolves and is optimized over the course of a few days. With this approach, the company has improved click-through rates by as much as 1900% (*14*). The challenge in extending such implicit algorithms to other Web applications will be to convert user interactions into a fitness assignment.

As interactive evolutionary algorithms improve and are adopted by Web site developers, we expect them to become increasingly useful for adding intelligence to interactive Web sites. Web sites with explicit interactive evolutionary algorithms could allow users to custom-design products by interactively browsing through virtual catalogs that evolve as users surf through them. Implicit algorithms could enable search engines to adaptively improve their responses to search queries over time and produce user-customized responses. This intelligent Web of the future will not just be powered by better algorithms, but will emerge from the interactions of millions of online users.

**References and Notes**
1. K. A. De Jong, *Evolutionary Computation: A Unified Approach* (MIT Press, Cambridge, MA, 2006).
2. T. Wajima, M. Matsumoto, S. Sekino, *Hitachi Rev.* **54**, 161 (2005).
3. J. D. Lohn, G. S. Hornby, D. S. Linden, *Artif. Intell. Eng. Des. Anal. Manuf.* **22**, 235 (2008).
4. R. Dawkins, *The Blind Watchmaker* (Norton, New York, 1986).
5. K. Sims, *SIGGRAPH 91 Conference Proceedings, Annual Conference Series* (ACM Press, New York, 1991), pp. 319–328.
6. S. Todd, W. Latham, *Evolutionary Art and Computers* (Academic Press, Orlando, FL, 1992).
7. J. Frazer, *An Evolutionary Architecture* (Architectural Association Publications, London, 1995).
8. J. A. Biles, *Proceedings of the International Computer Music Conference* (ACMA, San Francisco, 1994), pp. 131–137.
9. M. Witbrock, S. Neil-Reilly in *Evolutionary Design by Computers*, P. J. Bentley, Ed. (Morgan Kaufmann, San Francisco, 1999), chap. 10.
10. H. Takagi, *Proc. IEEE* **89**, 1275 (2001).
11. A parameterized representation is like a blueprint in which the parameter values specify dimensions; for example, a table has the parameters of height, width, and depth of the top, plus additional parameters for the dimensions of the legs. A generative representation is a more expressive type of representation and—in contrast to the parameterized representation—allows for the topology of a design to be changed.
12. S. Wannarumon, E. L. J. Bohez, K. Annanon, *Artif. Intell. Eng. Des. Anal. Manuf.* **22**, 19 (2008).
13. T. Kurtoglu, M. I. Campbell, *Res. Eng. Des.* **20**, 59 (2009).
14. See www.snapads.com/casestudies.php.
15. Supported in part by NSF Creative-IT grant 0757532.

---

PHYSICS

# Edge-State Physics Without Magnetic Fields

A novel class of materials called topological insulators allows spin physics to be probed without the need for magnetic fields.

**Markus Büttiker**

Solids can be divided into conductors and insulators. A new class of materials, called topological insulators, has been predicted (*1*, *2*) that exhibit surface states that lead to quantized conductance of charge and spin. These surface states are helical edge states, which interconnect spin and momentum of the carriers. Observation of these states should not require application of a magnetic field. On page 294 of this issue, Roth *et al.* (*3*) present compelling experimental evidence for such helical edge states at the surface of a topological insulator—in this case, quantum wells of mercury telluride (HgTe). Related effects are seen in the quantum Hall effect, but only in the presence of high magnetic fields. In the quantum Hall effect, a magnetic field induces cyclotron motion of electrons that is essential for the formation of edge states.

In the band picture of solids, conduction in materials depends on where the chemical potential $\mu$ falls. In metals, it lies in the conduction band, but in insulators, it is at a lower energy and falls into the band gap between the valence and conduction bands. Topological insulators (*1*, *2*) are band insulators with particular symmetry properties arising from spin-orbit interactions. According to theory, the surface edge states should reflect the nontrivial topological properties of the band structure, leading to unidirectional carrier motion along the sample boundary. For the system to be time-invariant, the states must come in pairs so that along a horizontal bar of the material, each edge has a set of states allowing propagation to the left, and another allowing propagation to the right (see the figure, panel A).

At equilibrium, when both states of the pairs are equally populated, there is no net charge current. However, carriers moving to the right could all have spins pointing up and carriers moving to the left might all have spins pointing down. This situation would lead to a net circulating spin current that persists at equilibrium. Such a robust effect normally requires an applied magnetic field.

When a potential is applied, a net carrier flow is set up through a non-equilibrium population of edge states. In panel A of the figure, the flow is to the right, and the greater population of states is depicted with a thicker line.

The closest analog to this system is the quantum Hall effect in a two-dimensional electron gas, in which the bulk is insulating but the sample edge has chiral states that describe electron motion along the sample boundaries. The situation is particularly simple in the integer quantum Hall effect, where each edge has an integer number of states, all of which carry charge in the same direction independent of the spin directions

Department of Theoretical Physics, University of Geneva, 24 Quai E. Ansermet, 1211 Geneva, Switzerland. E-mail: markus.buttiker@unige.ch
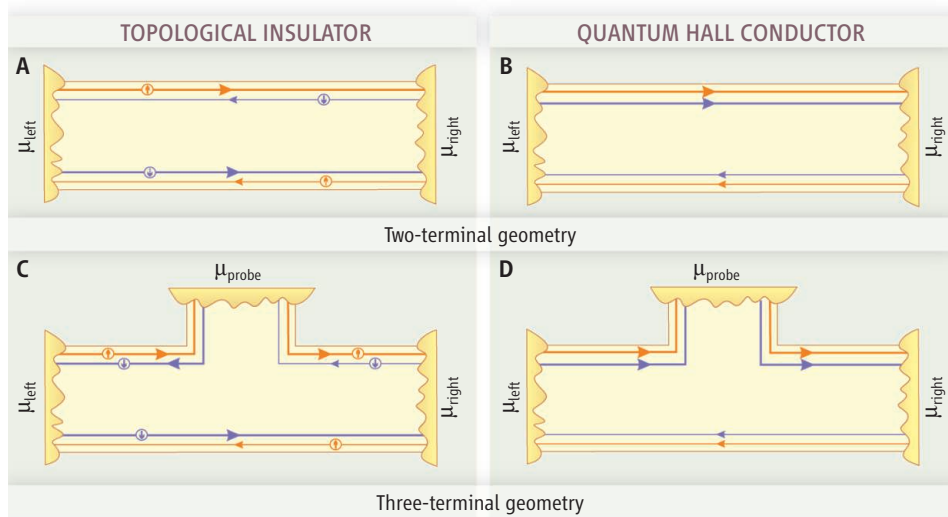
**Different at the edges.** Both topological insulators (**A** and **C**) and quantum Hall conductors (**B** and **D**) have edge states that connect adjacent contacts. For the topological insulator in (A), the edge states are in pairs that allow spin transport (red, spin up; blue, spin down). In all these examples, the higher chemical potential for the left contact ($\mu_{left}$) leads to transport to the right. In the presence of transport, thick lines show greater occupancy of the states, versus the thin lines showing the unoccupied channels. The topological insulator has states with high occupancy on both sides of the sample, whereas in the quantum Hall conductor, states with higher occupation occur only on one side (B). Adding a probe with chemical potential ($\mu_{probe}$) has no effect for the quantum Hall conductor in (D), but for the topological insulator, a current flows back from the probe to the left contact, cutting the overall conductance by half a conductance quantum (C).

but determined by the applied magnetic field (see the figure, panel B). This system has the same conductance as the topological insulator, but the two edge states that transport current in the integer quantum Hall effect are on the same (upper) side of the sample. In contrast, in the topological insulator, there is one edge state on each side of the sample (see the figure, panel A).

Molenkamp and co-workers provided experimental evidence for topological insulators (*4*) by measuring a conductance consistent with the theoretically predicted number of edge states. However, it remained to be shown that the observed conductance measurements arose from edge states. Roth *et al.* now show changes in conductance caused by a voltage probe. The theoretical predictions are based on a simple application of the scattering theory of multiprobe conductance (*5*) to topological insulators with multiple contacts.

The main effect can be understood by considering the conductance of a three-probe conductor, with one contact playing the role of a voltage probe (*1*). At such a contact, the net current vanishes. Electrons that leave the conductor are replaced by electrons from the contact reservoir. In the quantum Hall effect sample, two edge states from the left source contact enter the voltage probe, and two edge states leave the probe to the right drain contact (see the figure, panel D). The potential of the probe is equal to that of the source con-

tact, and the voltage probe has no effect on the overall conductance.

The situation is very different for a topological insulator (see the figure, panel C). Here, only one edge state is directed from the source contact to the voltage probe. Two other edge states lead away from the probe—one to the source contact and one to the sink contact. To maintain zero current, it is sufficient to tune the chemical potential at the probe halfway between the potentials of the source and sink contact. Now, half the current is directed back to the source contact (a channel that was not populated in the two-terminal geometry in panel A). The voltage probe reduces the overall conductance by half a conductance quantum. Roth *et al.* observed this effect in several different multiprobe geometries.

In the topological insulator, a voltage probe that maintains zero charge current provides momentum relaxation, because it forces half the carriers back against their direction of incidence. Such a probe maintains zero net charge current into the contact. However, the spin current into the probe is nonzero and net spin up in the case depicted. Simultaneously, a spin current is induced into both the source and drain electrodes.

Conceptually, we could ask for a probe that nullifies both the charge and the spin current. It would require separate potentials for spin-up and spin-down carriers. It would, in analogy to the quantum Hall effect, have

no effect on conductance. The spin-up potential would be equal to the source contact and the spin-down potential equal to the sink contact. Only one edge state would be filled along both the upper and lower edges. Experiments that relied on unequally populated edge states (*5*) were used almost 20 years ago to prove the physical reality of edge states in the quantum Hall effect (*6*–*8*).

Much of the physics of topological insulators is still under active investigation. Weak disorder has no effect, but strong disorder has been found theoretically to lead to a new phase (*9*) termed a "topological Anderson insulating state." The material used by Roth *et al.* is not the only topological insulator being explored. Martin *et al.* have proposed (*10*) that applying gates to bilayer graphene should generate edge states that are wide relative to the underlying lattice constants of graphene. Topological states may also exist at the interface of ferromagnets and superconductors deposited on top of a topological insulator (*11, 12*). Such heterostructures could be used to generate qubits (the working states of quantum computers) that are largely immune to the limiting effects of decoherence.

Edge states can be used to direct electrons from one place to another, in a manner similar to directing beams of photons with optics. Such capabilities are of interest in quantum information and quantum processing. The first steps in this direction will be the realization in topological insulators of basic building blocks such as quantum point contacts, Mach-Zehnder interferometers, and the demonstration of two-particle effects (*13, 14*) without the use of a magnetic field.

**References**
1. C. L. Kane, E. J. Mele, *Phys. Rev. Lett.* **95**, 226801 (2005).
2. B. A. Bernevig, T. L. Hughes, S.-C. Zhang, *Science* **314**, 1757 (2006).
3. A. Roth *et al.*, *Science* **325**, 294 (2009).
4. M. König *et al.*, *Science* **318**, 766 (2007); published online 19 September 2007 (10.1126/science.1148047).
5. M. Büttiker, *Phys. Rev. B* **38**, 9375 (1988).
6. S. Komiyama *et al.*, *Phys. Rev.* B **40**, 12566 (1989).
7. B. J. van Wees *et al.*, *Phys. Rev. Lett.* **62**, 1181 (1989).
8. B. W. Alphenaar *et al.*, *Phys. Rev. Lett.* **64**, 677 (1990).
9. J. Li, R.-L. Chu, J. K. Jain, S.-Q. Shen, *Phys. Rev. Lett.* **102**, 136806 (2009).
10. I. Martin, Ya. M. Blanter, A. F. Morpurgo, *Phys. Rev. Lett.* **100**, 036804 (2008).
11. L. Fu, C. L. Kane, *Phys. Rev. Lett.* **102**, 216403 (2009).
12. A. R. Akhmerov, J. Nilsson, C. W. J. Beenakker, *Phys. Rev. Lett.* **102**, 216404 (2009).
13. P. Samuelsson, E. V. Sukhorukov, M. Büttiker, *Phys. Rev. Lett.* **92**, 026805 (2004).
14. I. Neder *et al.*, *Nature* **448**, 333 (2007).

# Dyslexia: A New Synergy Between Education and Cognitive Neuroscience

John D. E. Gabrieli

Reading is essential in modern societies, but many children have dyslexia, a difficulty in learning to read. Dyslexia often arises from impaired phonological awareness, the auditory analysis of spoken language that relates the sounds of language to print. Behavioral remediation, especially at a young age, is effective for many, but not all, children. Neuroimaging in children with dyslexia has revealed reduced engagement of the left temporo-parietal cortex for phonological processing of print, altered white-matter connectivity, and functional plasticity associated with effective intervention. Behavioral and brain measures identify infants and young children at risk for dyslexia, and preventive intervention is often effective. A combination of evidence-based teaching practices and cognitive neuroscience measures could prevent dyslexia from occurring in the majority of children who would otherwise develop dyslexia.

The nexus of educational policies, evidence-based teaching practices, and cognitive neuroscience promises to use cutting-edge scientific methods and concepts to promote the growth and success of children. Reading is a focal point in this new synthesis because it is the most important portal to knowledge in our information age, from books to blackboards to the Internet. Learning to read is, however, perilous for the 5 to 17% of children who have developmental dyslexia, a persistent difficulty in learning to read that is not explained by sensory deficits, cognitive deficits, lack of motivation, or lack of adequate reading instruction (1).

Here, I provide an overview of research about the cognitive and brain bases of dyslexia, its treatment and brain plasticity associated with successful treatment, and how neuroscience may interact with education to help children with dyslexia. Particularly promising is the possibility that early identification of risk for dyslexia, through combined behavioral and neuroscience measures, may allow for preventive treatment such that many children with dyslexia who would otherwise fail to read would, instead, succeed at reading.

principles to acquire accurate and fluent reading skills, (ii) reduced vocabulary and strategies needed for text comprehension, and (iii) reduced motivation to read. The latter reasons for reading failure often involve socioeconomic factors, at home and at school, that are beyond the scope of this review.

An initial difficulty in learning to read has wide and prolonged consequences. Difficulty in reading discourages children with dyslexia to practice their reading outside of the classroom, and lack of practice alone can impede the growth of reading skill and the acquisition

reading comprehension in the later years of education, as students shift from learning to read to reading to learn.

Dyslexia appears on a continuum with typical reading ability because specific psychological, neural, and genetic features of dyslexia also correlate with reading performance in a broad range of children. On one hand, this means that dyslexia may be understood in terms of normative psychological and computational models of reading and that discoveries about dyslexia may offer insights into mechanisms of normal reading acquisition (3, 4). On the other hand, education and research findings depend on what behavioral boundary or criteria is selected to operationally define dyslexia. Dyslexia is often defined by a discrepancy between an average or above-average score on a test of general intelligence [intelligence quotient (IQ) test] and a low score on a standardized reading test. The core mechanism of dyslexia, however, appears to be similar in dyslexic readers, regardless of IQ over a broad range of IQ scores such that that children with low reading and IQ scores benefit from the same treatments as children with discrepant scores (5). These findings are consistent with the observation that dyslexia is independent of other talents that allow some children with dyslexia to grow into remarkably successful adults.

Dyslexia is strongly (54 to 75%) heritable, occurring in up to 68% of identical twins and 50% of individuals who have a parent or sibling with dyslexia (6). Environmental factors are also
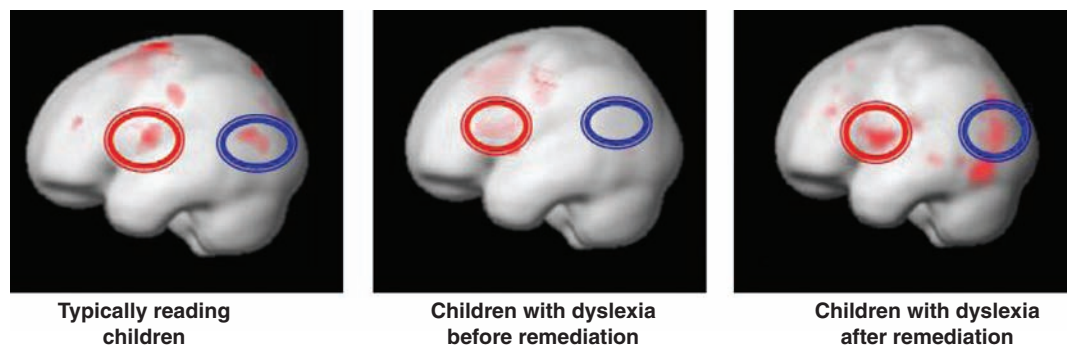


**Fig. 1.** Brain activation differences in dyslexia and its treatment [from (36)]. Functional magnetic resonance imaging activations shown on the left hemisphere for phonological processing in typically developing readers (**left**), age-matched dyslexic readers (**middle**), and the difference before and after remediation in the same dyslexic readers (**right**). Red circles identify the frontal region, and blue circles identify the temporo-parietal region of the brain. Both regions are hypoactivated in dyslexia and become more activated after remediation.

Below Fig. 1, left to right: **Typically reading children** · **Children with dyslexia before remediation** · **Children with dyslexia after remediation**

## What Is Dyslexia and What Causes Dyslexia?

*Definition of dyslexia.* Most children have reading difficulties for three broad reasons: (i) dyslexia, which is characterized by a difficulty in understanding and using alphabetic or logographic

Department of Brain and Cognitive Sciences and Harvard–Massachusetts Institute of Technology (MIT) Division of Health Sciences and Technology and McGovern Institute for Brain Research, MIT, 43 Vassar Street, Cambridge, MA 02139, USA. E-mail: gabrieli@mit.edu

of vocabulary and world knowledge (2). There are massive reading practice differences between good and poor readers: Outside of school in 5th grade, a good reader may read as many words in 2 days as a poor reader does in an entire year. Dyslexia is persistent: A student who fails to read adequately in 1st grade has a 90% probability of reading poorly in 4th grade and a 75% probability of reading poorly in high school. Thus, difficulty in early reading limits

important in reading development, even in children at genetic risk for dyslexia. For example, heritability is greater among children whose parents have a higher educational level (7). This suggests that genetic risk factors account for more variance in highly supportive environments, but less so in environments that vary widely in support for reading. Identified candidate risk genes (8) are implicated in neural migration and brain development, which suggests that dyslexia may

be a consequence of atypical neural migration in the developing brain.

*Psychological bases of dyslexia.* The causes of dyslexia can be considered at multiple levels of analysis and probably reflect multiple interacting mechanisms that vary across children. Historically, dyslexia was termed "word blindness"; however, the most common psychological cause of dyslexia for English speakers is a deficit in auditory processing of the sounds of language (phonological processing) (*9*). The diagnosis of dyslexia in the United States is commonly made in children ages 7 to 8 years old when reading difficulty is clearly measurable, although there is consensus that the roots of dyslexia begin before initial reading instruction, around 6 years of age (1st grade).

Beginning readers must decode print to access the identity and meaning of words. They already know the meanings of words in spoken language, but they have to learn to relate language to print through explicit phonological awareness that spoken words are composed of discrete sounds (phonemes) that can be mapped onto letters or syllables (graphemes). Children with dyslexia frequently exhibit poor phonological awareness, initially for spoken words and subsequently for printed words. These children have difficulty performing oral tasks that depend on phonological awareness, such as deciding which words start with the same sound as "hat"—"bat," "hot," or "sun," segmenting words into parts (knowing that "hat" is composed of "h," "a," and "t" sounds, or that those separate sounds can be blended into "hat"), and selectively deleting a sound within a word [what word remains if you take the "l" sound out of "clap" ("cap")]. For older children who can read, phonological impairment is most evident when asked to read aloud nonsense words ("twale") that are unknown and can only be pronounced or decoded on the basis of grapheme-to-phoneme mapping principles. These problems in phonological processing result in inaccurate recognition of words.

The expression of phonological difficulty in dyslexia varies as a consequence of differences in written languages (orthographies) (*10*). In alphabetic languages, such as English and Spanish, letters correspond to speech sounds, whereas in logographic languages, such as Chinese, characters correspond to meanings (morphemes). Alphabetic languages vary in their regularity (how consistently letters or letter clusters relate to one speech sound). Spanish and Italian are far more regular than English. Cross-cultural studies have shown that learning to read single words (grapheme-phoneme decoding) takes longer in less consistent orthographies. Current research suggests that across languages there are similar rates of dyslexia and that weakness in phonological processing is the most common etiology of dyslexia, but that the precise relation of phonological processing to reading and to the expression of dyslexia may vary across orthographies.

The second major problem for many children with dyslexia involves fluent reading of text. Even children who improve their accuracy for reading single words often continue to read text laboriously and slowly; the effort expended to read words in text often detracts from their ability to construct the meaning of what they are reading. This dysfluency may reflect a slowness evident even for naming a series of objects or colors. Children who have difficulties in both phonology and speed are described as having a double deficit (*11*). The dysfluency may also reflect difficulties in making up for the enormous amount of reading practice that these students miss out on when they remain poor readers in middle or late elementary grades (*12*). Much less is understood about the fluency deficit than the phonological deficit in dyslexia, but the fluency deficit is problematic for older children who must read increasingly sophisticated texts.

Scientists have been interested in discovering whether broader perceptual deficits precede reading deficits in dyslexia. Perhaps because these perceptual processes are less directly measurable in relation to reading and may exert their influences early in language development, there is debate about their precise role in dyslexia. The rapid temporal processing hypothesis derives from studies of children with "specific language impairment," a developmental language disorder estimated to occur in 7% of preschool children; these children have a difficulty in phonological awareness and/or morphosyntax, and they often progress to having dyslexia (*13*). Many of these children perform poorly at identifying the order of rapidly presented tones (*14*), and it is hypothesized that a broad auditory temporal processing deficit compromises accurate discrimination of language sounds that depends on very brief differences in auditory inputs (e.g., "b" and "d" differ by 50 msec or less of auditory information). The "magnocellular hypothesis" (*15*) is motivated by postmortem evidence in dyslexia for reduced area of the magnocellular layers of the lateral geniculate nucleus of the thalamus (*16*), which is part of the pathway mediating transient visual percepts such as motion. Individuals with dyslexia have exhibited subtle deficits in processing rapidly changing visual nonverbal information (e.g., gratings) and correlations between degrees of such visual impairment and reading difficulty (*17*). Other researchers report that children with dyslexia have, instead, a perceptual deficit in the exclusion of visual or auditory noise (*18*, *19*) or deficient stimulus-specific adaptation mechanisms (*20*). Conflicting reports on the presence or relation of these perceptual deficits to dyslexia raise the possibility that the relation between broader perceptual difficulties and reading difficulty may vary across children with dyslexia.

*Brain basis of dyslexia.* Functional neuroimaging studies have revealed differences in brain function and connectivity that are characteristic of dyslexia. Specific patterns of atypical brain activation in dyslexia relate to the specific reading or language processes examined in a neuroimaging study. When performing tasks that demand phonological awareness for print, such as deciding whether or not letters, words, or pseudoword letter strings rhyme, typically developing child and adult readers recruit several brain regions, including the left temporoparietal cortex. In contrast, children and adults with dyslexia exhibit reduced or absent activation in this region (Fig. 1) (*21*–*23*). Hypoactivation of the left temporo-parietal cortex is evident when dyslexic children are compared with typically developing readers who are three years younger and reading at the same level as the dyslexic children (*24*). Therefore, left temporo-parietal hypoactivation appears to be related to the etiology of dyslexia per se, rather than delayed maturation or reading level. It is hypothesized that this left temporo-parietal region supports the cross-modal relation of auditory and visual processes during reading. Atypical activations in dyslexia are also found in the left prefrontal regions associated with verbal working memory [in some cases related to reading ability rather than dyslexia (*24*)], left middle and superior temporal gyri associated with receptive language, and left occipito-temporal regions associated with visual analysis of letters and words.

Functional neuroimaging studies have also examined cultural and perceptual influences on dyslexia. Adults with dyslexia in French, Italian, and English exhibit similar hypoactivation in the left temporal cortex (*25*). Chinese readers with dyslexia exhibit atypical activation in the left prefrontal cortex, but not in the left temporo-parietal regions that are commonly atypical in dyslexic individuals reading alphabetic languages (*26*). Dyslexic children do not show activation during the incidental auditory perception of rapidly (relative to slowly) changing non-speech stimuli that is shown in the left prefrontal cortex by typically developing children, but dyslexic children do show increased activation after remediation with a computer-based program focused on improving rapid auditory processing (*27*). There is also reduced or absent activation in individuals with dyslexia in response to gratings designed to preferentially stimulate the magnocellular pathway in visual cortices (*28*, *29*). Further, reading ability correlates with individual differences in activation in response to these nonverbal visual stimuli (*29*). Also, contrast responsivity to nonverbal stimuli in the motion-sensitive visual cortex correlates with behavioral measures of phonological awareness in children with a wide range of reading skills (*30*).

White-matter pathways of the brain may be characterized by diffusion tensor imaging (DTI), which provides a quantitative index of the organization of large myelinated axons constituting the long-range connections of brain networks. White-matter organization appears to be weaker in the left posterior brain region of people with dyslexia than is typical (*31*), and this measure of organization correlates positively with reading scores among both typical and
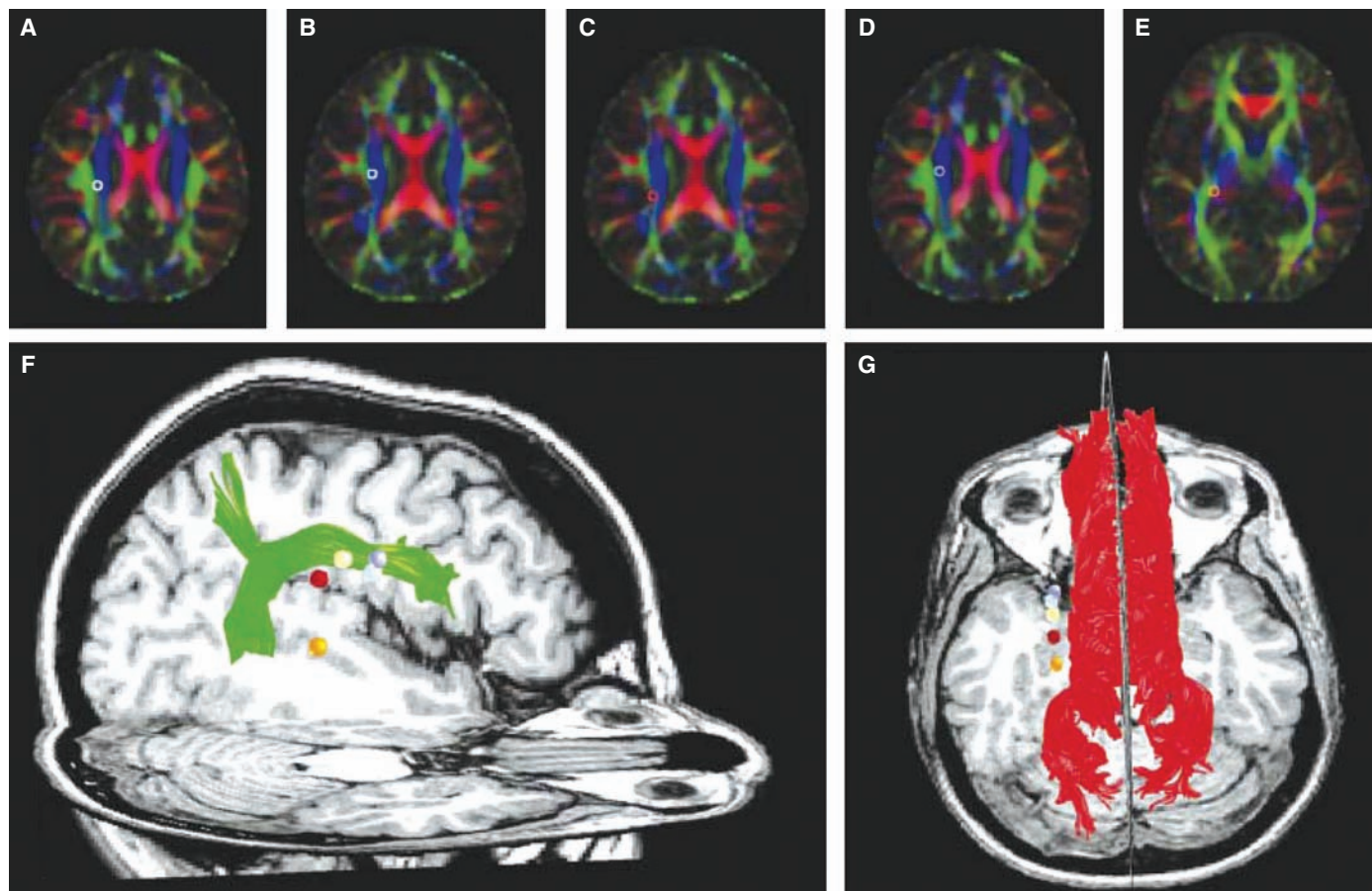
**Fig. 2.** Reading-related group differences in white matter as measured by DTI [from (48)]. Top row (**A** to **E**) shows reading-related differences in five independent studies; same locations of group differences are viewed sagitally (**F**) and axially (**G**). Colors correspond to estimated directions of white-matter pathways: left-right, red; anterior-posterior, green; inferior-superior, blue.

dyslexic readers (Fig. 2) (31–33). DTI studies of dyslexia also report greater-than-normal white-matter connectivity in the corpus callosum, the large white matter tract connecting homotopic regions of the left and right hemispheres (34). These findings suggest that, in dyslexia, white-matter pathways supporting reading project too weakly within the primary reading pathways of the linguistic left hemisphere, but they project too strongly between hemispheres (which may reflect an atypical reliance on right-hemisphere regions for reading that is observed in a number of functional neuroimaging studies). DTI is suitable for young children because its measurement does not require task performance. Studies with children conducted before reading instruction may determine whether the differential organization of white matter is predictive of developing dyslexia or is a consequence of reading practice.

### Can Dyslexia Be Treated?

*Remediation of dyslexia.* Once children are diagnosed with dyslexia because of reading failure, treatments are instructional. Typical public school and special education interventions often stabilize the degree of reading failure rather than remediate (normalize) reading skill (35). Well-controlled studies involving random assignment to treatment and control groups consistently show

that instruction yields substantial improvement in reading accuracy for many, but not all, children if instruction is more intensive (for instance, 100 min per day for 8 weeks), occurs in small groups (1 or 2 students per teacher), and includes explicit and systematic instruction in phonological awareness and decoding strategies (although the proportion of such instruction relative to reading meaningful text can vary widely with similar success). Gains are maintained for at least a year or two by ~50% of children after they return to the school's standard curriculum. Those children who retain their benefits improve from year to year, but they do not further catch up to typical readers. Such improvements are much more likely to occur in children who are beginning to read (ages 6 to 8) than in older children and are much more difficult to achieve for fluency than for accuracy. Thus, these resource-demanding interventions are effective for many children, but there are still challenges in developing interventions that are effective for all children.

*How remediation of dyslexia alters the brain.* Functional neuroimaging studies have revealed brain plasticity associated with effective intervention for dyslexia. In general, effective remediation is associated with increased activation, or normalization, in the left temporo-parietal and frontal regions that typically show reduced or

absent activation in dyslexia for phonological processing of visually presented letters, words, or sentences (36–40). Immediately after intervention, increased right-hemisphere activations are also observed (36–39). Typical reading development is characterized by decreased right-hemisphere engagement and increased left-hemisphere engagement (41), which may reflect a shift in interpreting visual inputs like letters and words from specific percepts to categorical linguistic representations. Thus, individuals with dyslexia receiving intervention may engage, in a contracted period, both right- and left-hemisphere mechanisms underlying reading development. These changes in brain function can be maintained for at least a year after remediation is completed and students have returned to their standard curriculum (37, 40).

Neuroimaging studies have not yet revealed what is different in the brains of children who do or do not respond to an intervention or sustain the benefits of intervention. It would be especially useful if neuroimaging markers were identified that could predict, before a specific intervention is provided, which children would benefit from a treatment, so that a given child could be offered an intervention most likely to help that child. To be informative, such neuroimaging studies would need to be longitudinal and involve many participants so that variation

among children with dyslexia could be characterized rigorously.

## Can Dyslexia Be Predicted and Prevented?

A major goal for all behavioral disorders is their prevention. Dyslexia is currently identified by reading failure that is difficult for the child and that discourages reading practice. If children at risk for dyslexia could be identified before reading instruction or early during this process (between infancy and 1st grade), there is opportunity to intervene therapeutically and minimize or eliminate reading failure.

There is good evidence that dyslexia can be predicted and prevented in many children. Individually administered screening assessments for children in kindergarten and 1st grade have been developed that are brief and easy to give and yield strong predictions about future reading ability; these assessments focus on knowledge of letter names and sounds, phonological awareness, and speed of naming. Further, when beginning readers identified as "at risk" are provided with the sort of intensive instruction described above, 56 to 92% of at-risk children across six studies were brought within the range of average reading ability (42). Further, early intervention reduces the risk of the difficult-to-remediate fluency deficit that emerges in 4th grade.

One challenge regards the specificity of screening measures. It is estimated that to identify all of the weakest 10% of beginning readers, current measures would identify 20% of children as being at high risk. Because effective prevention is resource-demanding, more accurate identification of at-risk children would be valuable.

*Brain measures predict risk for language and reading difficulty.* Longitudinal studies have shown that brain measures can predict future language and reading problems in infants and young children before reading instruction. These studies measured event-related potentials (ERPs), which are time-locked changes in electrical activity in response to stimuli measured with scalp electrodes that have excellent temporal (millisecond) resolution, although the brain locations of the sources of the electrical activity are uncertain. ERPs can be performed readily with infants and children, so that brain mechanisms relevant for ultimate language and reading achievement can be measured before overt manifestations of language or reading. Most of these studies examined infants and children with familial risk for reading disorders to have a reasonably large percentage of participants go on to exhibit reading difficulties.

Newborns from families with versus without familial risk for dyslexia exhibit differences in ERP responses to language sounds within hours or days of birth, a finding all the more impressive because only about half the newborns with familial risk are expected to become dyslexic years later (43). Longitudinal ERP studies have shown impressive relations between brain responses at infancy and later language and reading success or failure. ERP responses to speech

sounds within 36 hours of birth discriminated with over 81% accuracy those infants who would go on to become dyslexic readers at age 8 (44). Newborns, tested within a week of birth, had ERPs in response to speech sounds that correlated with language scores at ages 2.5, 3.5, and 5 years of age (45). These studies indicate that brain differences are present near the time of birth that greatly enhance the risk for and underscore the developmental nature of dyslexia. The findings also suggest that a deep understanding of the developmental pathways that lead to dyslexia demand prospective, longitudinal studies, from birth to early reading experience around ages 6 to 8.

Perhaps the most practical, near-term synergy between education and cognitive neuroscience arises from an integration of behavioral and brain measures in the service of predicting reading difficulty and then offering intervention to avoid reading failure. One example of this synergy comes from a study focused on decoding, the ability to determine the sound of a letter string from its constituent letters and syllables (46). Children identified by teachers as being at risk for reading difficulty at the start of a school year received a standardized test of decoding and 12 additional behavioral measures of language and reading, and they also underwent brain imaging. The behavioral and brain measures taken at the beginning of the school year were then related to the children's decoding ability at the end of the same school year, which improved on average after a year of education. The behavioral test scores and the brain imaging values in the fall accounted for 65 and 57%, respectively, of the variance in end-of-year decoding performance, but the combination of behavioral and brain measures accounted for significantly more of the variance (81%). Another longitudinal study related ERP measures in kindergarten to reading performance 5 years later and found that the addition of the ERP measures not only improved the prediction of reading ability over behavioral measures alone, but that only the ERP measures significantly predicted reading success in 5th grade (47). In both studies, brain measures significantly enhanced accuracy, beyond that possible with behavioral measures alone, in predicting long-term reading outcomes in children.

These findings suggest that the combination of behavioral and brain measures, perhaps together with genetic and familial information, may enhance the certainty with which dyslexia can be predicted for a child and promote the possibility of preventive intervention that allows many more children to succeed at learning to read.

## References and Notes
1. S. Shaywitz, *Overcoming Dyslexia* (Vintage Books, New York, 2003).
2. A. E. Cunningham, K. E. Stanovich, *Am. Educ.* **22**, 8 (1998).
3. K. Rayner, B. R. Foorman, C. A. Perfetti, D. Pesetsky, M. S. Seidenberg, *Psychol. Sci. Public Interest* **2**, 31 (2001).
4. J. C. Ziegler et al., *Cognition* **107**, 151 (2008).
5. K. E. Stanovich, *Learn. Disabil. Q.* **28**, 103 (2005).
6. B. Pennington, J. Gilger, in *Developmental Dyslexia: Neural, Cognitive, and Genetic Mechanisms*, C. H. Chase, G. D. Rosen, G. F. Sherman, Eds. (York, Baltimore, 1996), pp. 41–61.
7. A. Friend, J. C. DeFries, R. K. Olson, *Psychol. Sci.* **19**, 1124 (2008).
8. A. M. Galaburda, J. LoTurco, F. Ramus, R. H. Fitch, G. D. Rosen, *Nat. Neurosci.* **9**, 1213 (2006).
9. L. Bradley, P. E. Bryant, *Nature* **271**, 746 (1978).
10. J. C. Ziegler, U. Goswami, *Psychol. Bull.* **131**, 3 (2005).
11. M. Wolf, P. G. Bowers, *J. Educ. Psychol.* **91**, 415 (1999).
12. J. K. Torgesen, C. A. Rashotte, A. Alexander, in *Dyslexia, Fluency, and the Brain*, M. Wolf, Ed. (York, Timonium, MD, 2001), pp. 333–355.
13. J. B. Tomblin et al., *J. Speech Lang. Hear. Res.* **40**, 1245 (1997).
14. P. Tallal, M. Piercy, *Nature* **241**, 468 (1973).
15. J. Stein, *Dyslexia* **7**, 12 (2001).
16. M. S. Livingstone, G. D. Rosen, F. W. Drislane, A. M. Galaburda, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7943 (1991).
17. J. B. Demb, G. M. Boynton, M. Best, D. J. Heeger, *Vision Res.* **38**, 1555 (1998).
18. A. J. Sperling, Z. Lu, F. R. Manis, M. S. Seidenberg, *Nat. Neurosci.* **8**, 862 (2005).
19. J. C. Ziegler, C. Pech-Georgel, F. George, C. Lorenzi, *Dev. Sci.*, in press (10.1111/j.1467-7687.2009.00819.x).
20. M. Ahissar, Y. Lubin, H. Putter-Katz, K. Banai, *Nat. Neurosci.* **9**, 1558 (2006).
21. J. M. Rumsey et al., *Arch. Neurol.* **49**, 527 (1992).
22. S. E. Shaywitz et al., *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2636 (1998).
23. E. Temple et al., *Neuroreport* **12**, 299 (2001).
24. F. Hoeft et al., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4234 (2007).
25. E. Paulesu et al., *Science* **291**, 2165 (2001).
26. W. T. Siok, C. A. Perfetti, Z. Jin, L. H. Tan, *Nature* **431**, 71 (2004).
27. N. Gaab, J. D. Gabrieli, G. K. Deutsch, P. Tallal, E. Temple, *Restor. Neurol. Neurosci.* **25**, 295 (2007).
28. G. F. Eden et al., *Nature* **382**, 66 (1996).
29. J. B. Demb, G. M. Boynton, D. J. Heeger, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13363 (1997).
30. M. Ben-Shachar, R. F. Dougherty, G. K. Deutsch, B. A. Wandell, *Neuroimage* **37**, 1396 (2007).
31. T. Klingberg et al., *Neuron* **25**, 493 (2000).
32. G. K. Deutsch et al., *Cortex* **41**, 354 (2005).
33. C. Steinbrink et al., *Neuropsychologia* **46**, 3170 (2008).
34. R. F. Dougherty et al., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8556 (2007).
35. J. K. Torgesen, in *The Science of Reading: A Handbook*, M. Snowling, C. Hulme, Eds. (Blackwell, Malden, MA, 2006), chap. 27.
36. E. Temple et al., *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2860 (2003).
37. B. A. Shaywitz et al., *Biol. Psychiatr.* **55**, 926 (2004).
38. E. H. Aylward et al., *Neurology* **61**, 212 (2003).
39. G. F. Eden et al., *Neuron* **44**, 411 (2004).
40. A. Meyler, T. A. Keller, V. L. Cherkassky, J. D. Gabrieli, M. A. Just, *Neuropsychologia* **46**, 2580 (2008).
41. P. E. Turkeltaub, L. Gareau, D. L. Flowers, T. A. Zeffiro, G. F. Eden, *Nat. Neurosci.* **6**, 767 (2003).
42. J. K. Torgesen, *Am. Educ.* **Fall**, 6 (2004).
43. T. K. Guttorm, P. H. Leppanen, U. Richardson, H. Lyytinen, *J. Learn. Disabil.* **34**, 534 (2001).
44. D. L. Molfese, *Brain Lang.* **72**, 238 (2000).
45. T. K. Guttorm et al., *Cortex* **41**, 291 (2005).
46. F. Hoeft et al., *Behav. Neurosci.* **121**, 602 (2007).
47. U. Maurer et al., *Biol. Psychiatr.*, in press (10.1016/j.biopsych.2009.02.031).
48. M. Ben-Shachar, R. F. Dougherty, B. A. Wandell, *Curr. Opin. Neurobiol.* **17**, 258 (2007).
49. This work was supported by the Ellison Medical Foundation, MIT Class of 1976 Funds for Dyslexia Research, and B. Richmond and J. Richmond through the Martin Richmond Memorial Fund. I thank J. Torgesen, P. Hook, D. Willingham, I. Christodoulou, I. Kovelman, T. Perrachione, S. Whitfield-Gabrieli, and C. Gabrieli for comments on the paper and P. O'Loughlin and J. Gabrieli for help with the manuscript.

10.1126/science.1171999

# Foundations for a New Science of Learning

Andrew N. Meltzoff,[1,2,3]* Patricia K. Kuhl,[1,3,4] Javier Movellan,[5,6] Terrence J. Sejnowski[5,6,7,8]

Human learning is distinguished by the range and complexity of skills that can be learned and the degree of abstraction that can be achieved compared with those of other species. *Homo sapiens* is also the only species that has developed formal ways to enhance learning: teachers, schools, and curricula. Human infants have an intense interest in people and their behavior and possess powerful implicit learning mechanisms that are affected by social interaction. Neuroscientists are beginning to understand the brain mechanisms underlying learning and how shared brain systems for perception and action support social learning. Machine learning algorithms are being developed that allow robots and computers to learn autonomously. New insights from many different fields are converging to create a new science of learning that may transform educational practices.

Cultural evolution, which is rare among species and reaches a pinnacle in *Homo sapiens*, became possible when new forms of learning evolved under selective pressure in our ancestors. Culture underpins achievements in language, arts, and science that are unprecedented in nature. The origin of human intelligence is still a deep mystery. However, the study of child development, the plasticity of the human brain, and computational approaches to learning are laying the foundation for a new science of learning that provides insights into the origins of human intelligence.

Human learning and cultural evolution are supported by a paradoxical biological adaptation: We are born immature. Young infants cannot speak, walk, use tools, or take the perspective of others. Immaturity comes at a tremendous cost, both to the newborn, whose brain consumes 60% of its entire energy budget (*1*), and to the parents. During the first year of life, the brain of an infant is teeming with structural activity as neurons grow in size and complexity and trillions of new connections are formed between them. The brain continues to grow during childhood and reaches the adult size around puberty. The development of the cerebral cortex has "sensitive periods" during which connections between neurons are more plastic and susceptible to environmental influence: The sensitive periods for sensory processing areas occur early in development, higher cortical areas mature later, and the prefrontal cortex continues to develop into early adulthood (*2*).

Yet immaturity has value. Delaying the maturation and growth of brain circuits allows initial learning to influence the developing neural architecture in ways that support later, more complex learning. In computer simulations, starting the learning process with a low-resolution sensory system allows more efficient learning than starting with a fully developed sensory system (*3*).

What characterizes the exuberant learning that occurs during childhood? Three principles are emerging from cross-disciplinary work in psychology, neuroscience, machine learning, and education, contributing to a new science of learning (Fig. 1). These principles support learning across a range of areas and ages and are particularly useful in explaining children's rapid learning in two unique domains of human intelligence: language and social understanding.

*Learning is computational.* Discoveries in developmental psychology and in machine learning are converging on new computational accounts of learning. Recent findings show that infants and young children possess powerful computational skills that allow them automatically to infer structured models of their environment from the statistical patterns they experience. Infants use statistical patterns gleaned from experience to learn about both language and causation. Before they are three, children use frequency distributions to learn which phonetic units distinguish words in their native language (*4, 5*), use the transitional probabilities between syllables to segment words (*6*), and use covariation to infer cause-effect relationships in the physical world (*7*).

Machine learning has the goal of developing computer algorithms and robots that improve automatically from experience (*8, 9*). For example, BabyBot, a baby doll instrumented with a video camera, a microphone, and a loudspeaker (*10*), learned to detect human faces using the temporal contingency between BabyBots' programmed vocalizations and humans that tended to

[1]Institute for Learning and Brain Sciences, University of Washington, Seattle, WA 98195, USA. [2]Department of Psychology, University of Washington, Seattle, WA, 98195, USA. [3]Learning in Informal and Formal Environments (LIFE) Center, University of Washington, Seattle, WA, 98195, USA. [4]Department of Speech and Hearing Sciences, University of Washington, Seattle, WA, 98195, USA. [5]Institute for Neural Computation, University of California at San Diego, La Jolla, CA 92093, USA. [6]Temporal Dynamics of Learning Center (TDLC), University of California at San Diego, La Jolla, CA 92093, USA. [7]Howard Hughes Medical Institute, Salk Institute for Biological Studies, La Jolla, CA 92037, USA. [8]Division of Biological Sciences, University of California at San Diego, La Jolla, CA 92093, USA.

*To whom correspondence should be addressed. E-mail: Meltzoff@u.washington.edu

**Fig. 1.** The new science of learning has arisen from several disciplines. Researchers in developmental psychology have identified social factors that are essential for learning. Powerful learning algorithms from machine learning have demonstrated that contingencies in the environment are a rich source of information about social cues. Neuroscientists have found brain systems involved in social interactions and mechanisms for synaptic plasticity that contribute to learning. Classrooms are laboratories for discovering effective teaching practices. [Photo credits: R. Goebel (neuroscience), iStockphoto.com/J. Bryson (education), Y. Tsuno/AFP/Getty Images (machine learning)]

respond to these babylike vocalizations. After 6 min of learning, BabyBot detected novel faces and generalized to the schematic faces used in studies of infant face recognition.

Statistical regularities and covariations in the world thus provide a richer source of information than previously thought. Infants' pickup of this information is implicit; it occurs without parental training and begins before infants can manipulate the physical world or speak their first words. New machine learning programs also succeed without direct reinforcement or supervision. Learning from probabilistic input provides an alternative to Skinnerian reinforcement learning and Chomskian nativist accounts (11, 12).

*Learning is social.* Children do not compute statistics indiscriminately. Social cues highlight what and when to learn. Even young infants are predisposed to attend to people and are motivated to copy the actions they see others do (13). They more readily learn and reenact an event when it is produced by a person than by an inanimate device (14, 15).

Machine learning studies show that systematically increasing a robot's social-like behaviors and contingent responsivity elevates young children's willingness to connect with and learn from it (16). Animal models may help explain how social interaction affects learning: In birds, neurosteroids that affect learning modulate brain activity during social interaction (17). Social interaction can extend the sensitive period for learning in birds (18). Social factors also play a role in lifelong learning—new social technologies (for example, text messaging, Facebook, and Twitter) tap humans' drive for social communication. Educational technology is increasingly embodying the principles of social interaction in intelligent tutoring systems to enhance student learning (19).

*Learning is supported by brain circuits linking perception and action.* Human social and language learning are supported by neural-cognitive systems that link the actions of self and other. Moreover, the brain machinery needed to perceive the world and move our bodies to adapt to the movements of people and objects is complex, requiring continuous adaptation and plasticity. Consider what is necessary to explain human imitative learning. Newborns as young as 42 min old match gestures shown to them, including tongue protrusion and mouth opening (20). This is remarkable because infants cannot see their own faces, and newborns have never seen their reflection in a mirror. Yet, newborns can map from observed behavior to their own matching acts, suggesting shared representations for the acts of self and others (15, 20). Neuroscientists have discovered a striking overlap in the brain systems recruited both for the perception and production of actions (21, 22). For example, in human adults there is neuronal activation when observing articulatory movements in the cortical areas responsible for producing those articulations (23). Social learning, imitation, and sensorimotor experience may initially generate, as well as modify and refine, shared neural circuitry for perception and action. The emerging field of social neuroscience is aimed at discovering brain mechanisms supporting close coupling and attunement between the self and other, which is the hallmark of seamless social communication and interaction.

## Social Learning and Understanding

Human children readily learn through social interactions with other people. Three social skills are foundational to human development and rare in other animals: imitation, shared attention, and empathic understanding.

*Imitation.* Learning by observing and imitating experts in the culture is a powerful social learning mechanism. Children imitate a diverse range of acts, including parental mannerisms, speech patterns, and the use of instruments to get things done. For example, a toddler may see her father using a telephone or computer keyboard and crawl up on the chair and babble into the receiver or poke the keys. Such behavior is not explicitly trained (it may be discouraged by the parent), and there is no inborn tendency to treat plastic boxes in this way—the child learns by watching others and imitating.

Imitation accelerates learning and multiplies learning opportunities. It is faster than individual discovery and safer than trial-and-error learning. Children can use third-person information (observation of others) to create first-person knowledge. This is an accelerator for learning: Instead of having to work out causal relations themselves, children can learn from watching experts. Imitative learning is valuable because the behavioral actions of others "like me" serve as a proxy for one's own (15).

Children do not slavishly duplicate what they see but reenact a person's goals and intentions. For example, suppose an adult tries to pull apart an object but his hand slips off the ends. Even at 18 months of age, infants can use the pattern of unsuccessful attempts to infer the unseen goal of another. They produce the goal that the adult was striving to achieve, not the unsuccessful attempts (14). Children choose whom, when, and what to imitate and seamlessly mix imitation and self-discovery to solve novel problems (24, 25).

Imitation is a challenging computational problem that is being intensively studied in the robotic and machine learning communities (26, 27). It requires algorithms capable of inferring complex sensorimotor mappings that go beyond the repetition of observed movements. The match must be achieved despite the fact that the teacher may be different from the observer in several ways (e.g., size, spatial orientation, morphology, dexterity). The ultimate aim is to build robots that can learn like infants, through observation and imitation (28). Current computational approaches to imitation can be divided into direct and goal-based approaches. Direct approaches learn input-action mappings that reproduce the observed behaviors (26); goal-based approaches, which are more recent and less explored, infer the goal of the observed behaviors and then produce motor plans that achieve those goals (29, 30).

*Shared attention.* Social learning is facilitated when people share attention. Shared attention to the same object or event provides a common ground for communication and teaching. An early component of shared attention is gaze following (Fig. 2). Infants in the first half year of life look more often in the direction of an adult's head turn when peripheral targets are in the visual field (31). By 9 months of age, infants interacting with a responsive robot follow its head movements, and the timing and contingencies, not just the visual appearance of the robot, appear to be key (32). It is unclear, however, whether young infants are trying to look at what another is seeing or are simply tracking head movements. By 12 months, sensitivity to the direction and state of the eyes exists, not just sensitivity to the direction of head turning. If a person with eyes open turns to one of



**Fig. 2.** Gaze following is a mechanism that brings adults and infants into perceptual contact with the same objects and events in the world, facilitating word learning and social communication. After interacting with an adult (**left**), a 12-month-old infant sees an adult look at one of two identical objects (**middle**) and immediately follows her gaze (**right**).

two equidistant objects, 12-month-old infants look at that particular target, but not if the person makes the same head movement with eyes closed (*33*).

A blindfold covering the person's eyes causes 12-month-olds to make the mistake of following the head movements. They understand that eye closure, but not a blindfold, blocks the other person's view. Self-experience corrects this error. In a training study, 1-year-olds were given experience with a blindfold so they understood that it made it impossible to see. When the adult subsequently wore the blindfold, infants who had received self-experience with it treated the adult as if she could not see (*34*), whereas control infants did not. Infants project their own experience onto other people. The ability to interpret the behavior and experience of others by using oneself as a model is a highly effective learning strategy that may be unique to humans and impaired in children with autism (*35, 36*). It would be useful if this could be exploited in machine learning, and preliminary progress is being made (*37*).

*Empathy and social emotions.* The capacity to feel and regulate emotions is critical to understanding human intelligence and has become an active area of research in human-computer interaction (*38*). In humans, many affective processes are uniquely social. Controlled experiments lead to the conclusion that prelinguistic toddlers engage in altruistic, instrumental helping (*39*). Children also show primitive forms of empathy. When an adult appears to hurt a finger and cry in pain, children under 3 years of age comfort the adult, sometimes offering a bandage or teddy bear (*40*). Related behavior has been observed with children helping and comforting a social robot that was "crying" (*16, 41*).

Brain imaging studies in adults show an overlap in the neural systems activated when people receive a painful stimulus themselves or perceive that another person is in pain (*42, 43*). These neural reactions are modulated by cultural experience, training, and perceived similarity between self and other (*43, 44*). Atypical neural patterns have been documented in antisocial adolescents (*45*). Discovering the origins of individual differences in empathy and compassion is a key issue for developmental social-cognitive neuroscience.

## Language Learning

Human language acquisition poses a major challenge for theories of learning, and major advances have been made in the last decade (*46*). No computer has cracked the human speech code and achieved fluent speech understanding across talkers, which children master by 3 years of age (*11*).

Human language acquisition sheds light on the interaction among computational learning, social facilitation of learning, and shared neural circuitry for perception and production.

*Behavioral development.* Early in development, infants have a capacity to distinguish all sounds across the languages of the world, a capacity shared by nonhuman primates (*47*). However, infants' universal capacities narrow with development,
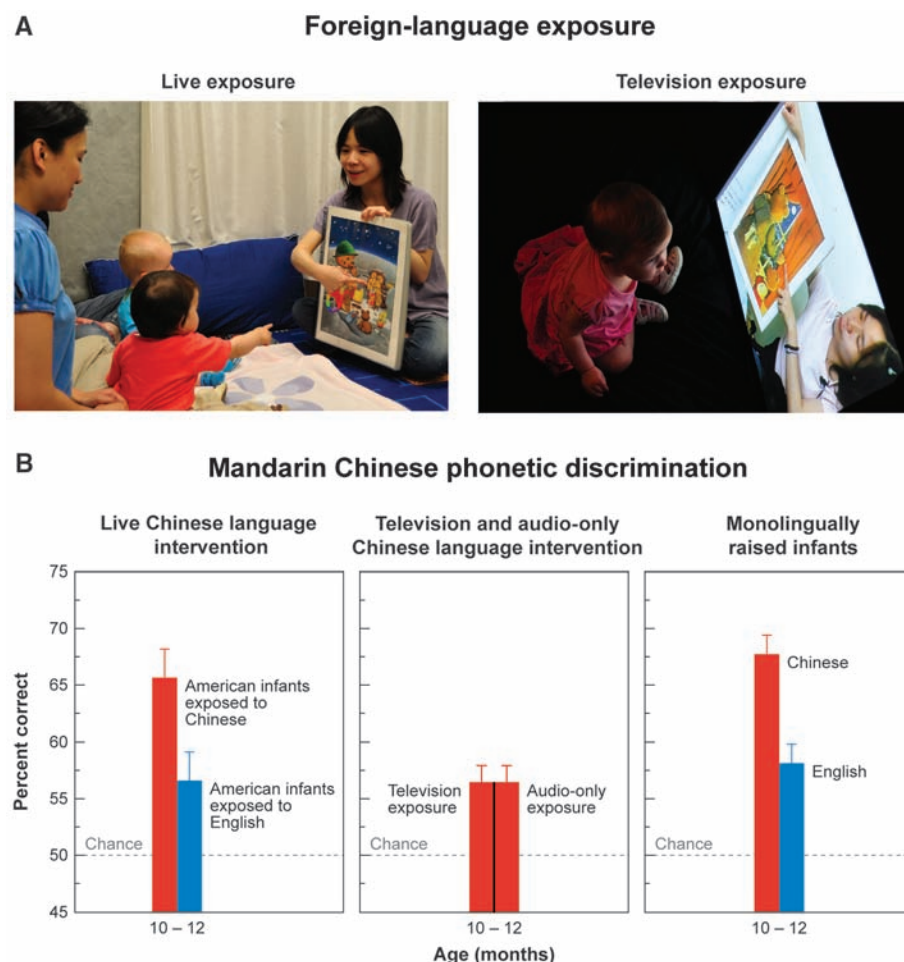


**Fig. 3.** The need for social interaction in language acquisition is shown by foreign-language learning experiments. Nine-month-old infants experienced 12 sessions of Mandarin Chinese through (**A**) natural interaction with a Chinese speaker (left) or the identical linguistic information delivered via television (right) or audiotape (not shown). (**B**) Natural interaction resulted in significant learning of Mandarin phonemes when compared with a control group who participated in interaction using English (left). No learning occurred from television or audiotaped presentations (middle). Data for age-matched Chinese and American infants learning their native languages are shown for comparison (right). [Adapted from (*50*) and reprinted with permission.]

and by one year of age, infants' ability to perceive sound distinctions used only in foreign languages and not their native environment is weakened. Infants' universal capacities become language-specific between 9 and 12 months of age. American and Japanese infants, who at 7 months of age discriminated /ra/ from /la/ equally well, both change by 11 months: American infants improve significantly while Japanese infants' skills show a sharp decline (*48*).

This transition in infant perception is strongly influenced by the distributional frequency of sounds contained in ambient language (*4, 5*). Infants' computational skills are sufficiently robust that laboratory exposure to artificial syllables in which the distributional frequencies are experimentally manipulated changes infants' abilities to discriminate the sounds.

However, experiments also show that the computations involved in language learning are "gated" by social processes (*49*). In foreign-language learning experiments, social interaction strongly influences infants' statistical learning. Infants exposed to a foreign language at 9 months learn rapidly, but only when experiencing the new language during social interchanges with other humans. American infants exposed in the laboratory to Mandarin Chinese rapidly learned phonemes and words from the foreign language, but only if exposed to the new language by a live human being during naturalistic play. Infants exposed to the same auditory input at the same age and for the same duration via television or audiotape showed no learning (*50*) (Fig. 3). Why infants learned better from people and what components of social interactivity support language learning are currently being investigated (*51*). Determining the key stimulus and interactive features will be important for theory. Temporal contingencies may be critical (*52*).

Other evidence that social input advances language learning comes from studies showing that infants vocally imitate adult vowel sounds by

5 months but not acoustically matched nonspeech sounds that are not perceived as human speech (*53, 54*). By 10 months, even before speaking words, the imitation of social models results in a change in the types of vocalizations children produce. Children raised in Beijing listening to Mandarin babble by using tonelike pitches characteristic of Mandarin, which make them sound distinctly Chinese. Children being raised in Seattle listening to English do not babble by using such tones and sound distinctly American.

Children react to a social audience by increasing the complexity of their vocal output. When mothers' responses to their infants' vocalizations are controlled experimentally, a mother's immediate social feedback results both in greater numbers and more mature, adultlike vocalizations from infants (*55*). Sensory impairments affect infant vocalizations: Children with hearing impairments use a greater preponderance of sounds (such as "ba") that they can see by following the lip movements of the talker. Infants who are blind babble by using a greater proportion of sounds that do not rely on visible articulations ("ga") (*56*).

Birdsong provides a neurobiological model of vocal learning that integrates self-generated sensorimotor experience and social input. Passerine birds learn conspecific song by listening to and imitating adult birds. Like humans, young birds listen to adult conspecific birds sing during a sensitive period in development and then practice that repertoire during a "sub-song" period (akin to babbling) until it is crystallized (*57*). Neural models of birdsong learning can account for this gradual process of successive refinement (*58*). In birds, as in humans, a social context enhances vocal learning (*59*).

*Neural plasticity.* In humans, a sensitive period exists between birth and 7 years of age when language is learned effortlessly; after puberty, new language learning is more difficult, and native-language levels are rarely achieved (*60, 61*). In



**Fig. 4.** A social robot can operate autonomously with children in a preschool setting. In this photo, toddlers play a game with the robot. One long-term goal is to engineer systems that test whether young children can learn a foreign language through interactions with a talking robot.

birds, the duration of the sensitive period is extended in richer social environments (*18, 62*). Human learning beyond the sensitive period may also benefit from social interaction. Adult foreign-language learning improves under more social learning conditions (*63*).

A candidate mechanism governing the sensitive period for language in humans is neural commitment (*11*). Neural commitment is the formation of neural architecture and circuitry dedicated to the detection of phonetic and prosodic characteristics of the particular ambient language(s) to which the infant is exposed. The neural circuitry maximizes detection of a particular language and, when fully developed, interferes with the acquisition of a new language.

Neural signatures of children's early language learning can be documented by using event-related potentials (ERPs). Phonetic learning can be documented at 11 months of age; responses to known words, at 14 months; and semantic and syntactic learning, at 2.5 years (*64*). Early mastery of the sound patterns of one's native language provides a foundation for later language learning: Children who show enhanced ERP responses to phonemes at 7.5 months show faster advancement in language acquisition between 14 and 30 months of age (*65*).

Children become both native-language listeners and speakers, and brain systems that link perception and action may help children achieve parity between the two systems. In adults, functional magnetic resonance imaging studies show that watching lip movements appropriate for speech activates the speech motor areas of the brain (*66*). Early formation of linked perception-production brain systems for speech has been investigated by using brain imaging technology called magnetoencephalography (MEG). MEG reveals nascent neural links between speech perception and production. At 6 months of age, listening to speech activates higher auditory brain areas (superior temporal), as expected, but also simultaneously activates Broca's area, which controls speech production, although listening to nonspeech sounds does not [(*67*); see also (*68*)]. MEG technology will allow linguists to explore how social interaction and sensorimotor experience affects the cortical processing of language in children and why young children can learn foreign language material from a human tutor but not a television.

New interactive robots are being designed to teach language to children in a social-like manner. Engineers created a social robot that autonomously interacts with toddlers, recognizing their moods and activities (*16*) (Fig. 4). Interaction with the social robot over a 10-day period resulted in a significant increase in vocabulary in 18- to 24-month-old children compared with the vocabulary of an age-matched control group (*41*). This robotic technology is now being used to test whether children might learn foreign language words through social games with the robot.

## Education

During their long period of immaturity, human brains are sculpted by implicit social and statistical learning. Children progress from relatively helpless, observant newborns to walking, talking, empathetic people who perform everyday experiments on cause and effect. Educators are turning to psychology, neuroscience, and machine learning to ask: Can the principles supporting early exuberant and effortless learning be applied to improve education?

Progress is being made in three areas: early intervention programs, learning outside of school, and formal education.

Children are born learning, and how much they learn depends on environmental input, both social and linguistic. Many children entering kindergarten in the United States are not ready for school (*69*), and children who start behind in school-entry academic skills tend to stay behind (*70*). Neuroscience work suggests that differences in learning opportunities before first grade are correlated with neural differences that may affect school learning (*71, 72*).

The recognition that the right input at the right time has cascading effects led to early interventions for children at risk for poor academic outcomes. Programs enhancing early social interactions and contingencies produce significant long-term improvements in academic achievement, social adjustment, and economic success and are highly cost effective (*73–75*).

The science of learning has also affected the design of interventions with children with disabilities. Speech perception requires the ability to perceive changes in the speech signal on the time scale of milliseconds, and neural mechanisms for plasticity in the developing brain are tuned to these signals. Behavioral and brain imaging experiments suggest that children with dyslexia have difficulties processing rapid auditory signals; computer programs that train the neural systems responsible for such processing are helping children with dyslexia improve language and literacy (*76*). The temporal "stretching" of acoustic distinctions that these programs use is reminiscent of infant-directed speech ("motherese") spoken to infants in natural interaction (*77*). Children with autism spectrum disorders (ASD) have deficits in imitative learning and gaze following (*78–80*). This cuts them off from the rich socially mediated learning opportunities available to typically developing children, with cascading developmental effects. Young children with ASD prefer an acoustically matched nonspeech signal over motherese, and the degree of preference predicts the degree of severity of their clinical autistic symptoms (*81*). Children

with ASD are attracted to humanoid robots with predictable interactivity, which is beginning to be used in diagnosis and interventions (*82*).

Elementary and secondary school educators are attempting to harness the intellectual curiosity and avid learning that occurs during natural social interaction. The emerging field of informal learning (*83*) is based on the idea that informal settings are venues for a significant amount of childhood learning. Children spend nearly 80% of their waking hours outside of school. They learn at home; in community centers; in clubs; through the Internet; at museums, zoos, and aquariums; and through digital media and gaming. Informal learning venues are often highly social and offer a form of mentoring, apprenticeship, and participation that maximizes motivation and engages the learner's sense of identity; learners come to think of themselves as good in technology or as budding scientists, and such self-concepts influence children's interests, goals, and future choices (*84, 85*). A recent National Research Council study on science education (*83*) cataloged factors that enliven learning in informal learning venues with the long-term goal of using them to enhance learning in school.

In formal school settings, research shows that individual face-to-face tutoring is the most effective form of instruction. Students taught by professional tutors one on one show achievement levels that are two standard deviations higher than those of students in conventional instruction (*86*). New learning technologies are being developed that embody key elements of individual human tutoring while avoiding its extraordinary financial cost. For example, learning researchers have developed intelligent tutoring systems based on cognitive psychology that provide an interactive environment with step-by-step feedback, feed-forward instructional hints to the user, and dynamic problem selection (*19*). These automatic tutors have been shown to approximate the benefits of human tutoring by adapting to the needs of individual students, as good teachers do. Classrooms are becoming living laboratories as researchers and educators use technology to track and collect data from individual children and use this information to test theories and design curricula.

## Conclusions

A convergence of discoveries in psychology, neuroscience, and machine learning has resulted in principles of human learning that are leading to changes in educational theory and the design of learning environments. Reciprocally, educational practice is leading to the design of new experimental work. A key component is the role of "the social" in learning. What makes social interaction such a powerful catalyst for learning? Can key elements be embodied in technology to improve learning? How can we capitalize on social factors to teach children better and to foster their natural curiosity about people and things? These are deep questions at the leading edge of the new science of learning.

## References and Notes

1. J. M. Allman, *Evolving Brains* (Freeman, New York, 1999).
2. S. R. Quartz, T. J. Sejnowski, *Behav. Brain Sci.* **20**, 537 (1997).
3. R. A. Jacobs, M. Dominguez, *Neural Comput.* **15**, 761 (2003).
4. P. K. Kuhl, K. A. Williams, F. Lacerda, K. N. Stevens, B. Lindblom, *Science* **255**, 606 (1992).
5. J. Maye, J. F. Werker, L. Gerken, *Cognition* **82**, B101 (2002).
6. J. R. Saffran, R. N. Aslin, E. L. Newport, *Science* **274**, 1926 (1996).
7. A. Gopnik *et al.*, *Psychol. Rev.* **111**, 3 (2004).
8. T. M. Mitchell, *Machine Learning* (McGraw-Hill, New York, 1997).
9. R. Douglas, T. Sejnowski, "Future challenges for the science and engineering of learning," www.nsf.gov/sbe/SLCWorkshopReportjan08.pdf.
10. N. J. Butko, I. R. Fasel, J. R. Movellan, in *Proceedings of the 5th IEEE International Conference on Development and Learning*, Bloomington, IN, 31 May to 3 June 2006.
11. P. K. Kuhl, *Nat. Rev. Neurosci.* **5**, 831 (2004).
12. A. Gopnik, J. B. Tannenbaum, *Dev. Sci.* **10**, 281 (2007).
13. A. N. Meltzoff, M. K. Moore, *Science* **198**, 75 (1977).
14. A. N. Meltzoff, *Dev. Psychol.* **31**, 838 (1995).
15. A. N. Meltzoff, *Acta Psychol.* **124**, 26 (2007).
16. F. Tanaka, A. Cicourel, J. R. Movellan, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17954 (2007).
17. L. Remage-Healey, N. T. Maidment, B. A. Schlinger, *Nat. Neurosci.* **11**, 1327 (2008).
18. M. S. Brainard, E. I. Knudsen, *J. Neurosci.* **18**, 3929 (1998).
19. K. R. Koedinger, V. Aleven, *Educ. Psychol. Rev.* **19**, 239 (2007).
20. A. N. Meltzoff, M. K. Moore, *Early Dev. Parenting* **6**, 179 (1997).
21. R. Hari, M. Kujala, *Physiol. Rev.* **89**, 453 (2009).
22. G. Rizzolatti, L. Fogassi, V. Gallese, *Nat. Rev. Neurosci.* **2**, 661 (2001).
23. R. Möttönen, J. Järveläinen, M. Sams, R. Hari, *Neuroimage* **24**, 731 (2004).
24. R. A. Williamson, A. N. Meltzoff, E. M. Markman, *Dev. Psychol.* **44**, 275 (2008).
25. B. M. Repacholi, A. N. Meltzoff, B. Olsen, *Dev. Psychol.* **44**, 561 (2008).
26. S. Schaal, *Trends Cogn. Sci.* **3**, 233 (1999).
27. A. P. Shon, J. J. Storz, A. N. Meltzoff, R. P. N. Rao, *Int. J. Hum. Robotics* **4**, 387 (2007).
28. Y. Demiris, A. N. Meltzoff, *Infant Child Dev.* **17**, 43 (2008).
29. A. Y. Ng, S. Russell, in *Proceedings of the 17th International Conference on Machine Learning*, Morgan Kaufmann, Stanford, CA, 29 June to 2 July 2000, pp. 663–670.
30. D. Verma, R. P. N. Rao, in *Advances in Neural Information Processing Systems* (MIT Press, Cambridge, MA, 2006), pp. 1393–1400.
31. R. Flom, K. Lee, D. Muir, Eds., *Gaze-Following* (Erlbaum, Mahwah, NJ, 2007).
32. J. R. Movellan, J. S. Watson, in *Proceedings of the 2nd IEEE International Conference on Development and Learning*, Cambridge, MA, 12 to 15 June 2002, pp. 34–42.
33. R. Brooks, A. N. Meltzoff, *Dev. Psychol.* **38**, 958 (2002).
34. A. N. Meltzoff, R. Brooks, *Dev. Psychol.* **44**, 1257 (2008).
35. A. N. Meltzoff, *Dev. Sci.* **10**, 126 (2007).
36. M. Tomasello, M. Carpenter, J. Call, T. Behne, H. Moll, *Behav. Brain Sci.* **28**, 675 (2005).
37. J. Bongard, V. Zykov, H. Lipson, *Science* **314**, 1118 (2006).
38. R. W. Picard, *Affective Computing* (MIT Press, Cambridge, MA, 1997).
39. F. Warneken, M. Tomasello, *Science* **311**, 1301 (2006).
40. C. Zahn-Waxler, M. Radke-Yarrow, R. A. King, *Child Dev.* **50**, 319 (1979).
41. J. R. Movellan, M. Eckhart, M. Virnes, A. Rodriguez, in *Proceedings of the International Conference on Human Robot Interaction*, San Diego, CA, 11 to 13 March 2009, pp. 307–308.
42. T. Singer *et al.*, *Science* **303**, 1157 (2004).
43. G. Hein, T. Singer, *Curr. Opin. Neurobiol.* **18**, 153 (2008).
44. C. Lamm, A. N. Meltzoff, J. Decety, *J. Cogn. Neurosci.*, in press; available online at www.mitpressjournals.org/doi/abs/10.1162/jocn.2009.21186.
45. J. Decety, K. J. Michalska, Y. Akitsuki, B. Lahey, *Biol. Psychol.* **80**, 203 (2009).
46. P. Kuhl, L. Gleitman, "Opportunities and challenges for language learning and education," www.nsf.gov/sbe/slc/NSFLanguageWorkshopReport.pdf.
47. P. K. Kuhl, J. D. Miller, *Science* **190**, 69 (1975).
48. P. K. Kuhl *et al.*, *Dev. Sci.* **9**, F13 (2006).
49. P. K. Kuhl, *Dev. Sci.* **10**, 110 (2007).
50. P. K. Kuhl, F.-M. Tsao, H.-M. Liu, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9096 (2003).
51. B. T. Conboy, P. K. Kuhl, in *On Being Moved: From Mirror Neurons to Empathy*, S. Bråten, Ed. (John Benjamins, Philadelphia, 2007), pp. 175–199.
52. P. R. Montague, T. J. Sejnowski, *Learn. Mem.* **1**, 1 (1994).
53. P. K. Kuhl, A. N. Meltzoff, *Science* **218**, 1138 (1982).
54. P. K. Kuhl, A. N. Meltzoff, *J. Acoust. Soc. Am.* **100**, 2425 (1996).
55. M. H. Goldstein, A. P. King, M. J. West, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8030 (2003).
56. C. Stoel-Gammon, in *Phonological Development*, C. A. Ferguson, L. Menn, C. Stoel-Gammon, Eds. (York, Timonium, MD, 1992), pp. 273–282.
57. M. S. Brainard, A. J. Doupe, *Nature* **417**, 351 (2002).
58. K. Doya, T. J. Sejnowski, in *Advances in Neural Information Processing Systems*, G. Tesauro, D. S. Touretzky, T. Leen, Eds. (MIT Press, Cambridge, MA, 1995), pp. 101–108.
59. A. J. Doupe, P. K. Kuhl, in *Neuroscience of Birdsong*, H. P. Zeigler, P. Marler, Eds. (Cambridge Univ. Press, Cambridge, 2008), pp. 5–31.
60. J. S. Johnson, E. L. Newport, *Cognit. Psychol.* **21**, 60 (1989).
61. R. I. Mayberry, E. Lock, *Brain Lang.* **87**, 369 (2003).
62. L. F. Baptista, L. Petrinovich, *Anim. Behav.* **34**, 1359 (1986).
63. Y. Zhang *et al.*, *Neuroimage* **46**, 226 (2009).
64. P. K. Kuhl, M. Rivera-Gaxiola, *Annu. Rev. Neurosci.* **31**, 511 (2008).
65. P. K. Kuhl *et al.*, *Philos. Trans. R. Soc. London Ser. B.* **363**, 979 (2008).
66. N. Nishitani, R. Hari, *Neuron* **36**, 1211 (2002).
67. T. Imada *et al.*, *Neuroreport* **17**, 957 (2006).
68. G. Dehaene-Lambertz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 14240 (2006).
69. J. P. Shonkoff, D. A. Philips, Eds., *From Neurons to Neighborhoods* (National Academy Press, Washington, DC, 2000).
70. G. J. Duncan *et al.*, *Dev. Psychol.* **43**, 1428 (2007).
71. R. D. S. Raizada, T. L. Richards, A. N. Meltzoff, P. K. Kuhl, *Neuroimage* **40**, 1392 (2008).
72. D. A. Hackman, M. J. Farah, *Trends Cogn. Sci.* **13**, 65 (2009).
73. C. T. Ramey, S. L. Ramey, *Merrill Palmer Q.* **50**, 471 (2004).
74. J. Heckman, *Science* **312**, 1900 (2006).
75. E. I. Knudsen, J. J. Heckman, J. L. Cameron, J. P. Shonkoff, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10155 (2006).
76. P. Tallal, *Nat. Rev. Neurosci.* **5**, 721 (2004).
77. P. K. Kuhl *et al.*, *Science* **277**, 684 (1997).
78. S. J. Rogers, H. G. Williams, Eds., *Imitation and the Social Mind: Autism and Typical Development* (Guilford, New York, 2006).
79. P. Mundy, L. Newell, *Curr. Dir. Psychol. Sci.* **16**, 269 (2007).
80. K. Toth, J. Munson, A. N. Meltzoff, G. Dawson, *J. Autism Dev. Disord.* **36**, 993 (2006).
81. P. K. Kuhl, S. Coffey-Corina, D. Padden, G. Dawson, *Dev. Sci.* **8**, F1 (2005).
82. B. Scassellati, in *Proceedings of the 14th IEEE International Workshop on Robot and Human Interactive Communication*, Nashville, TN, 13 to 15 August 2005, pp. 585–590.
83. P. Bell, B. Lewenstein, A. W. Shouse, M. A. Feder, Eds., *Learning Science in Informal Environments* (National Academy Press, Washington, DC, 2009).
84. C. D. Lee, *Educ. Res.* **37**, 267 (2008).
85. J. Bruner, *The Culture of Education* (Harvard Univ. Press, Cambridge, MA, 1996).
86. B. S. Bloom, *Educ. Res.* **13**, 4 (1984).
87. Supported by NSF Science of Learning Center grants SBE-0354453 (A.M. and P.K.) and SBE-0542013 (J.M. and T.S.), NIH grants HD-22514 (A.M.) and HD-37954 (P.K.), NSF IIS-0808767 (J.M.), and Howard Hughes Medical Institute (T.S.). The views in this article do not necessarily reflect those of NSF or NIH. We thank M. Asada for assistance with the robotic image in Fig. 1; A. Gopnik, J. Bongard, P. Marshall, S. Cheryan, P. Tallal, and J. Watson for valuable comments; and the members of the NSF Science of Learning Centers for helpful discussions.

# Modern Riding Style Improves Horse Racing Times

Thilo Pfau,* Andrew Spence, Sandra Starke, Marta Ferrari, Alan Wilson

Horse racing has existed in its current form for over 200 years. The apparently uncomfortable modern race riding posture was developed in the United States in the late 19th century, introduced to the United Kingdom in 1897, and universally adopted by 1910. This change in riding style (Fig. 1, A and B) corresponds to a dramatic improvement of 5 to 7% in race times in the United States between 1890 and 1900 and in the United Kingdom between 1897 and 1910 (fig. S1), strongly suggesting that the adopted posture benefits racing performance. This improvement is greater than that observed over the following century.

A jockey represents about 13% (~60 kg) of a horse's body mass (~450 kg), and both conventionally seated riders and sandbags elicit an increase in mechanical and metabolic cost proportionate to the mass of the load (1, 2). When weight is added to racehorses' saddlecloths to handicap them, a proportionate reduction in racing speed is observed (3). A

backpack frame that elastically rather than tightly couples the load to the wearer reduces the cost of load carrying, possibly attributed to a reduction in the vertical movement of the backpack and hence the potential energy changes that the wearer's legs must produce (4).

Changes in kinetic energy and gravitational potential energy of a horse during each stride of gallop are substantial (5) and may be associated with the

metabolic cost of galloping (6), despite energy-storing, springlike legs (7). We hypothesize that a jockey (Fig. 1B) uncouples himself from the horse by moving relative to his mount. Fast-running quadrupeds appear insensitive to increases in weight (effective gravity) (8), but increased inertia (with weight) is detrimental to athletic performance (1–3).

We measured acceleration and calculated displacement of horse (and jockey) by using Global Positioning System and two inertial sensors (9). Horse displacement was found to be $150 \pm 8$ mm vertically, and $100 \pm 7$ mm in the fore-aft direction; jockey

displacement was $60 \pm 9$ mm vertically, and $20 \pm 4$ mm in the fore-aft direction (Fig. 1C). The jockey's body moves little with respect to a world inertial frame, and therefore the horse supports the jockey's weight but does not have to accelerate and decelerate him or her through each stride cycle (Fig. 1C, top left). The jockey's legs oscillate in length while transmitting a vertical force fluctuating about the jockey's body weight, resulting in substantial mechanical work by the jockey (10). Interestingly, the jockey slightly overcompensates for the horse's motion (Fig. 1C). Thus, displacement and velocity (hence kinetic energy) fluctuations of the jockey plus horse system might be slightly smaller than that of the horse alone, and the jockey could possibly "drive" the horse.

The crouched posture may confer an additional small reduction in aerodynamic drag [≤130 W mechanical power (9); <2% horse mechanical work at gallop (8)], but the modern posture is high on the horse (Fig. 1B) so there is little change in frontal area and drag reduction is unlikely to be the primary goal or benefit (compare with a track cyclist's posture).



**Fig. 1. (A)** Jockeys demonstrating the traditional riding style in the United Kingdom (finish photo of 1900 Derby Stakes). [Reprinted from (11)] **(B)** The "martini glass" posture in modern thoroughbred racing (2008 Epsom Derby Stakes). [Credit: Tom Stanhope, Equine Action Images, www.equine-action-images.com] **(C)** Fore-aft (or craniocaudal, CC) and vertical (or dorsoventral, DV) movement of jockey (top left, green) and horse (bottom left, dark blue) with respect to constant velocity motion; movement of jockey relative to horse (middle right, light blue). Black lines indicate mean stride, and arrows, direction of movement. Phase in relation to contact of nonlead front leg (0%) is denoted by dots at 10% of stride values on the mean trajectory; air phase around 40 to 60%. Horses show larger amplitudes than jockeys, notably in the fore-aft direction. Horse and relative jockey movement are about 180° out of phase.

### References and Notes

1. C. R. Taylor, N. Heglund, T. A. McMahon, T. R. Looney, *J. Exp. Biol.* **86**, 9 (1980).
2. H. C. Schamhardt, H. W. Merkens, G. J. V. M. van Osch, in vol. 3 of *Equine Exercise Physiology*, S. G. B. Persson, A. Lindholm, L. B. Jeffcott, Eds. (International Conference on Equine Exercise Physiology, Davis, CA, 1991), pp. 120–127.
3. N. Mordin, *Winning Without Thinking: The Definitive Guide to Horse Race Betting Systems* (Aesculus, Derbyshire, UK, 2002), pp. 91–106.
4. L. C. Rome, L. Flynn, T. D. Yoo, *Nature* **444**, 1023 (2006).
5. T. Pfau, T. H. Witte, A. M. Wilson, *J. Exp. Biol.* **209**, 3742 (2006).
6. A. E. Minetti, L. P. Ardigo, E. Reinach, F. Saibene, *J. Exp. Biol.* **202**, 2329 (1999).
7. Y. H. Chang, R. J. Kram, *J. Appl. Physiol.* **86**, 1657 (1999).
8. J. R. Usherwood, A. M. Wilson, *Nature* **438**, 753 (2005).
9. Materials and methods are available as supporting material on *Science* Online.
10. E. A. Trowbridge, J. V. Cotterill, C. E. Crofts, *Eur. J. Appl. Physiol. Occup. Physiol.* **70**, 66 (1995).
11. V. Orchard, *The Derby Stakes* (Hutchinson, London, 1954).
12. We thank P. J. Hobbs for the horses; T. Cox, the Cox Library, and T. Stanhope, Equine Action Images, for photographs; the Horserace Betting Levy Board for funding (T.P., A.S., and M.F.); and the Bremer Ausbildungs Partnerschaft (international LEONARDO program) for funding (S.S.). The UK Biotechnology and Biological Sciences Research Council funded the technology development. A.W. holds a Royal Society Wolfson Research Merit award. A.W. and T.P. have filed a patent application on monitoring of horses and jockeys during racing with methods similar to those described here.

Structure and Motion Laboratory, The Royal Veterinary College, University of London, Hatfield, AL9 7TA UK.

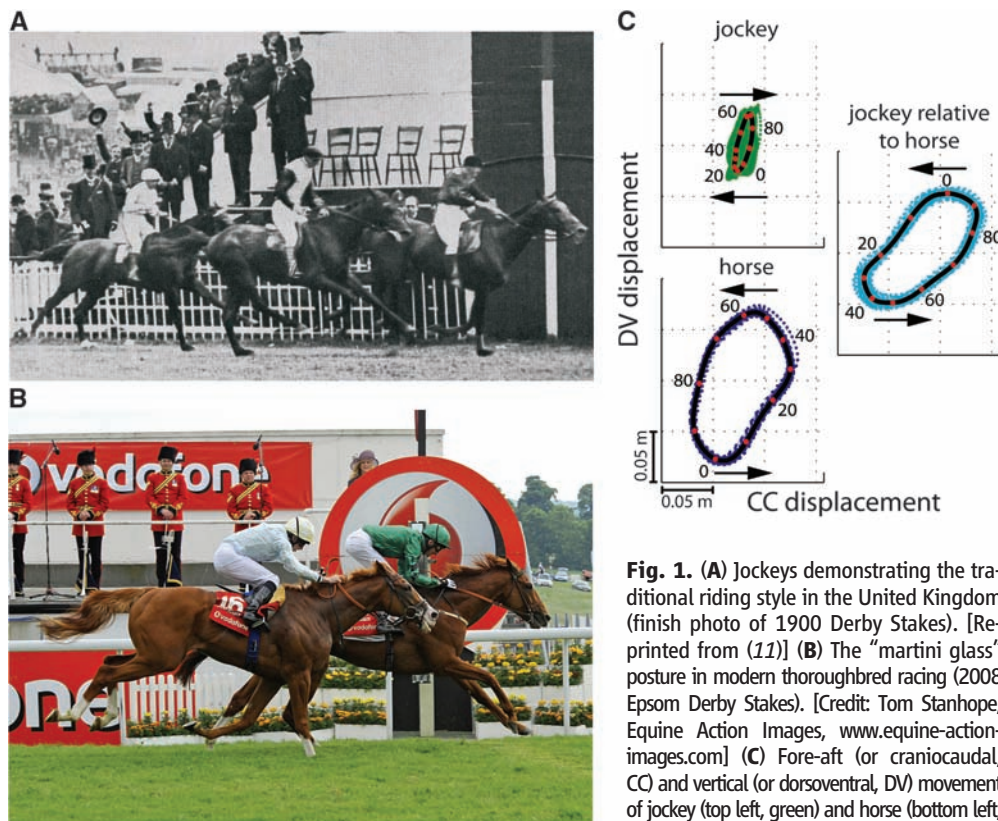*To whom correspondence should be addressed. E-mail: tpfau@rvc.ac.uk

# Demographic Variability, Vaccination, and the Spatiotemporal Dynamics of Rotavirus Epidemics

Virginia E. Pitzer,[1,2]* Cécile Viboud,[2] Lone Simonsen,[3] Claudia Steiner,[4]
Catherine A. Panozzo,[5] Wladimir J. Alonso,[2] Mark A. Miller,[2] Roger I. Glass,[2]
John W. Glasser,[5] Umesh D. Parashar,[5] Bryan T. Grenfell[1,2,6]

Historically, annual rotavirus activity in the United States has started in the southwest in late fall and ended in the northeast 3 months later; this trend has diminished in recent years. Traveling waves of infection or local environmental drivers cannot account for these patterns. A transmission model calibrated against epidemiological data shows that spatiotemporal variation in birth rate can explain the timing of rotavirus epidemics. The recent large-scale introduction of rotavirus vaccination provides a natural experiment to further test the impact of susceptible recruitment on disease dynamics. The model predicts a pattern of reduced and lagged epidemics postvaccination, closely matching the observed dynamics. Armed with this validated model, we explore the relative importance of direct and indirect protection, a key issue in determining the worldwide benefits of vaccination.

The importance of rotavirus as a leading cause of severe diarrhea among children in both developed and developing countries has captured the attention of policy-makers and vaccine manufacturers in recent years. Rota-virus is ubiquitous in the population; individuals are infected multiple times during their life span, usually via fecal-oral transmission (*1*). In adults, infection is typically asymptomatic; however, for children who have not been previously exposed to the virus, infection can lead to severe diarrhea and vomiting and may require hospitalization or result in death. About 610,000 deaths per year worldwide can be attributed to rotavirus (*2, 3*). In the United States, it is estimated that rotavirus annually leads to ~60,000 hospitalizations and 30 to 40 deaths among children <5 years of age (*4*). Two live, attenuated oral vaccines have recently been licensed in North America and Europe and have been demonstrated in clinical trials to prevent 79 to 98% of episodes of severe rotavirus diarrhea in vaccinated infants (*5–9*). Such vaccines have the potential to prevent considerable numbers of rotavirus-related hospitalizations and

[1]Center for Infectious Disease Dynamics, Pennsylvania State University, State College, PA 16801, USA. [2]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA. [3]School of Public Health and Health Services, George Washington University, Washington, DC 20052, USA. [4]Healthcare Cost and Utilization Project, Center for Delivery, Organization and Markets, Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services, Rockville, MD 20850, USA. [5]Epidemiology Branch, Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA. [6]Department of Ecology and Evolutionary Biology and Woodrow Wilson School, Princeton University, Princeton, NJ 08544, USA.

*To whom correspondence should be addressed. E-mail: vep2@psu.edu
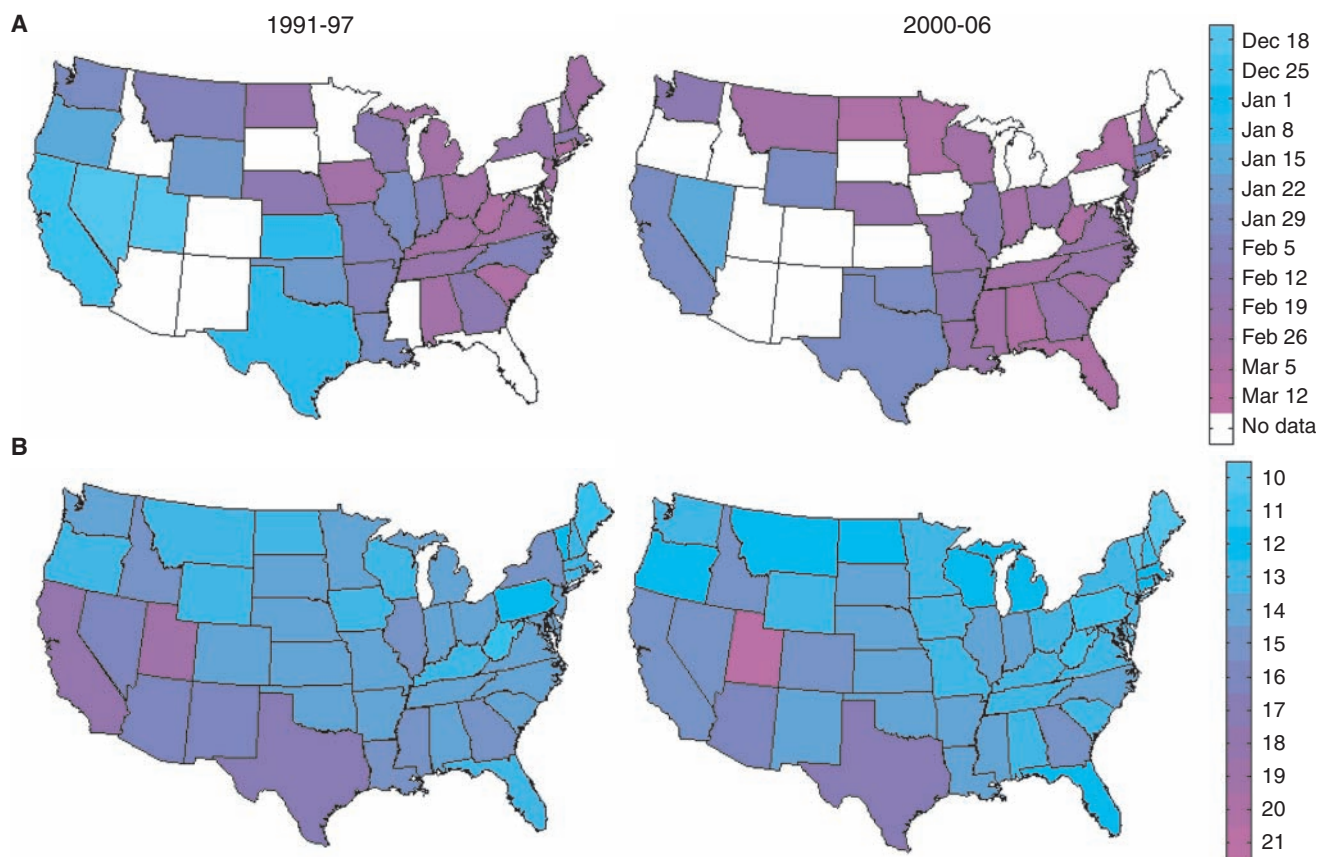
**Fig. 1.** Spatiotemporal pattern of rotavirus epidemics and birth rates in the United States. (**A**) Map of mean week of rotavirus activity for the period from 1991–1997 and 2000–2006. Mean timing of annual rotavirus activity was calculated for states that reported rotavirus-positive specimens to the NREVSS for at least four of the six rotavirus seasons. (**B**) Average crude birth rates by state (live births per 1000 estimated population per year) for 1991–1997 and 2000–2006 according to National Vital Statistics Reports.

deaths worldwide. Analyzing the impact of vaccination on the incidence of severe diarrhea, however, requires an understanding of the underlying dynamics of rotavirus infection.

Analysis of rotavirus epidemiological data from the continental United States indicates that disease incidence is highly seasonal and spatially structured. Until the late 1990s, annual epidemics peaked in the southwest in December to January and appeared to spread northeast, where epidemics peaked in March to April (Fig. 1A) (10–12). In recent years, this spatiotemporal pattern has become less pronounced, primarily because epidemics have occurred later in southwestern states (Fig. 1A) (13). The spatiotemporal pattern exhibited by rotavirus differs from that of other acute seasonal infections in the United States, notably influenza, for which epidemics tend to begin in major population centers relatively synchronously and spread to less populous areas in correlation with work commutes (14). Such hierarchical patterns are consistent with "gravity models" (15), which assume local extinction followed by annual reinvasion from outside of the United States (16). By contrast, the available epidemiological and viral phylogenetic evidence for rotavirus suggests local persistence of infection during epidemic troughs in summer months (10, 17–19).

If rotavirus does persist locally, then geographic differences in the timing of rotavirus epidemics must be driven by local heterogeneities in demographic or environmental factors. Although climatic factors have traditionally been suspected to be the underlying cause of rotavirus seasonality and patterns of spread (20), the spatiotemporal signature of rotavirus in the United States is too complex to be explained by a single environmental gradient (13). We quantified the impact of several environmental factors on rotavirus epidemic patterns, including solar radiation, precipitation, vapor pressure, and temperature. Univariate and multivariate regression suggested that such factors could not explain the observed variability in timing of rotavirus epidemics either across states or over time (10).

Alternatively, birth rates are known to play an important role in epidemic dynamics and can affect both regional spatial and temporal patterns (21–23). Birth rates have traditionally been ~30% higher in the southwestern United States than in the northeast, but over the past decade rates have declined by 10 to 25%, particularly in California (Fig. 1B) (24, 25). We found that the state-to-state and year-to-year variation in birth rates was consistent with the observed variability in the timing of rotavirus epidemics in the United States; that is, earlier epidemics tended to occur in localities having higher birth rates in the year leading up to the rotavirus season (Pearson's correlation coefficient $r = -0.57$, $P < 0.001$). Furthermore, regression analysis confirmed that birth rates were strongly associated with rotavirus epidemic timing, explaining 65 to 71% of the spatial variation across states and 61% of temporal trends (10) (table S1).

**Model description.** We developed an age- and state-stratified transmission model for rotavirus to test whether the spatiotemporal signature of epidemics could be generated by geographic variability in birth rate. The model allows for individuals to be infected multiple times over the course of their life; each infection is followed by a temporary period of complete immunity and confers an increased level of incomplete protection, both against infection and against severe diarrhea-given infection (Fig. 2A). This reflects the current understanding of rotavirus epidemiology and immunity (10).

The model assumes seasonal variation in the instantaneous transmission rate: $\beta(t) = \beta_0\{1 + a\cos[2\pi(t - \varphi)]\}$; here, $\beta_0$ is the baseline transmission rate, $a$ is the amplitude of seasonality, and $\varphi$ is a seasonal offset parameter. We used data from the State Inpatient Databases (SID) of the U.S. Healthcare Cost and Utilization Project (HCUP) (26) for a subset of 16 states with monthly hospital discharge data from January 1993 to December 2004. We explored four possible assumptions for mixing between age groups and found that the best-fitting model included an age-related peak in the rate of transmission to children <3 years of age [based on Akaike information criterion, (10)].

The rate of hospitalizations for rotavirus predicted by the model was similar to the reported incidence (Fig. 2B), as was the age distribution of
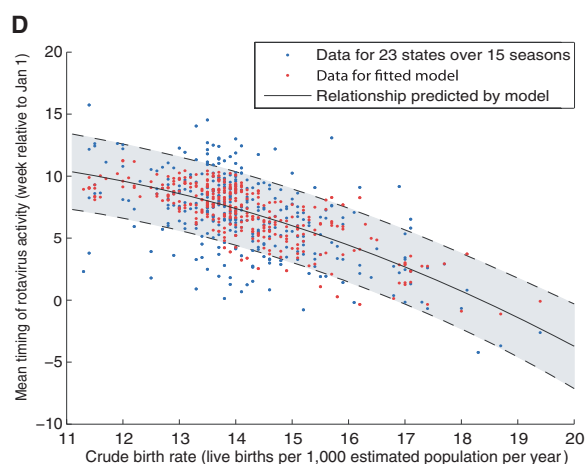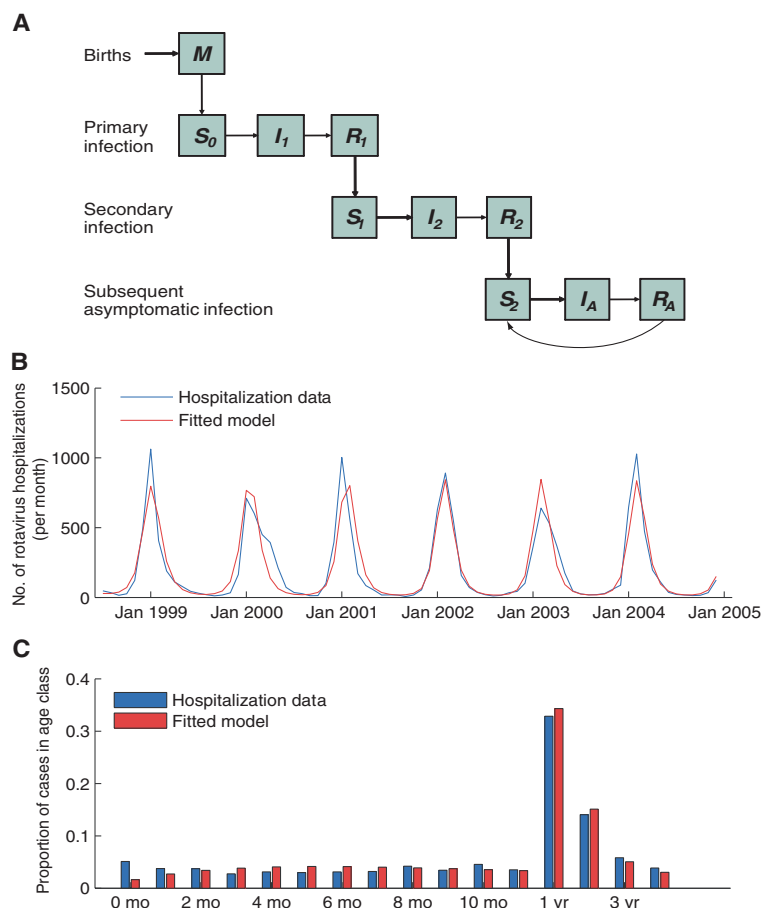


**Fig. 2.** Description of model and model fit. (**A**) Compartmental diagram illustrating the transmission model. See (10) for a more detailed description. (**B** and **C**) Fit of the model to age-specific hospitalization data from the California HCUP SID for (B) the number of hospitalizations per month in children 0 to 4 years of age and (C) the age distribution of hospitalized rotavirus cases (0 to 11 months and 1 to 4 years of age). (**D**) Timing of rotavirus epidemics based on laboratory-confirmed NREVSS data versus the state-specific crude birth rate in the preceding year. We calculated the mean timing of annual rotavirus activity for each of the 23 states and 15 rotavirus seasons (1991 to 2006) by weighting each calendar week by the proportion of rotavirus detections occurring that week. Observed data are plotted in blue, whereas the red dots represent predictions from the fitted model driven by birth rate. The relationship between the birth rate and timing of epidemics predicted by our model is given by the solid black line, whereas the shaded area between dashed lines represents the 95% confidence interval for the predicted relationship.

cases (<0.025 difference in the proportion of cases in any given 1-year age group) (Fig. 2C). When we incorporated information on the declining birth rate in California, the model predicted a seasonal peak of rotavirus activity gradually shifting from early January to February in this state (Fig. 2B). If we modeled a constant birth rate over this time period, the model was unable to capture the observed change in the timing of rotavirus epidemics and did not fit the data as well (10).

**Effect of birth rate on epidemic timing across the United States.** We then tested the birth rate effect for the entire country by using weekly data on the total number of rotavirus-positive specimens reported to the Centers of Disease Control and Prevention (CDC) through the National Respiratory and Enteric Virus Surveillance System (NREVSS). Laboratories throughout the United States voluntarily reported weekly information on rotavirus activity for the period from July (calendar week 27) 1991 to June (calendar week 26) 2006. We focused our analysis on the 23 states with at least 40 positive specimens reported each season over the study period.

The observed relationship between the timing of annual rotavirus activity (10) and mean birth rate was well captured by our fitted model (Fig. 2D). When the recruitment of susceptibles was high (around 20 live births per 1000 total population per year), epidemics tended to occur earlier (late November to early December), and when birth rate was low (around 12 per 1000), epidemics occurred on average 3 months later. A model with constant

average birth rate in space and time failed to capture the spatiotemporal dynamics of the disease (10).

**Impact of vaccination.** In February 2006, the RotaTeq vaccine (Merck) was introduced into the routine vaccination schedule for U.S. infants. Coverage with at least one dose of the vaccine by 3 months of age increased to about 60% by March 2008 (10). Data from NREVSS and the New Vaccine Surveillance Network (NVSN) indicate that the 2007–2008 rotavirus season was delayed by about 2 to 4 months and substantially diminished in size relative to epidemics in previous years (Fig. 3A) (27). Because vaccination effectively reduces the rate of recruitment of susceptible individuals, akin to a decline in birth rate, we can validate our model by comparing the predicted impact of vaccination to the observed dynamics.

We assumed that the vaccine conferred a protective effect similar to that of primary infection (average vaccine efficacy of 80.3% against severe diarrhea) (9). Although the vaccine does not completely protect against infection with rotavirus (vaccine efficacy of 36.5% against infection), it substantially decreases the probability of developing severe diarrhea (10). Weekly vaccine coverage levels were interpolated on the basis of monthly one-dose coverage estimates derived from health insurance claims data for children <3 months of age (10). We explored a variety of assumptions about the relative effectiveness of the one-dose coverage estimates compared with the proportion who receive the full three-dose protective effect conferred by the vaccine (10).

Our model predicted that the introduction of the RotaTeq vaccine in 2006 would lead to a small decrease in the incidence of severe diarrhea during the initial 2006–2007 season (Fig. 3B) and a larger decline and delay in the epidemic during 2007–2008 (Fig. 3C). These predicted dynamics captured the observed effects of vaccination (Fig. 3, A to C). If the relative effectiveness of the one-dose coverage estimates was assumed to be ~70%, as compared with the full three-dose vaccine schedule, then the model generated a good quantitative match with the observed impact of vaccination during 2007–2008 (Fig. 3, A to C). Model predictions suggest that the 2008–2009 season will be characterized by somewhat elevated rotavirus transmission during the summer and fall and a smaller epidemic peak occurring about 1 month later than in the prevaccination era. Thereafter, there will be a similar pattern of smaller yearly epidemics if vaccine coverage remains at its current level (~68% coverage with one dose) (Fig. 3D).

Lastly, we quantified the expected long-term impact of rotavirus vaccination at a variety of coverage levels. Vaccination is predicted to result in both decreased incidence of severe diarrhea and decreased prevalence of symptomatic and asymptomatic infections (Fig. 4). However, a single course of the vaccine will not eliminate the disease completely if the effect of vaccination is truly comparable to the protection provided by primary infection. This is unsurprising given that the duration of immunity against natural rotavirus
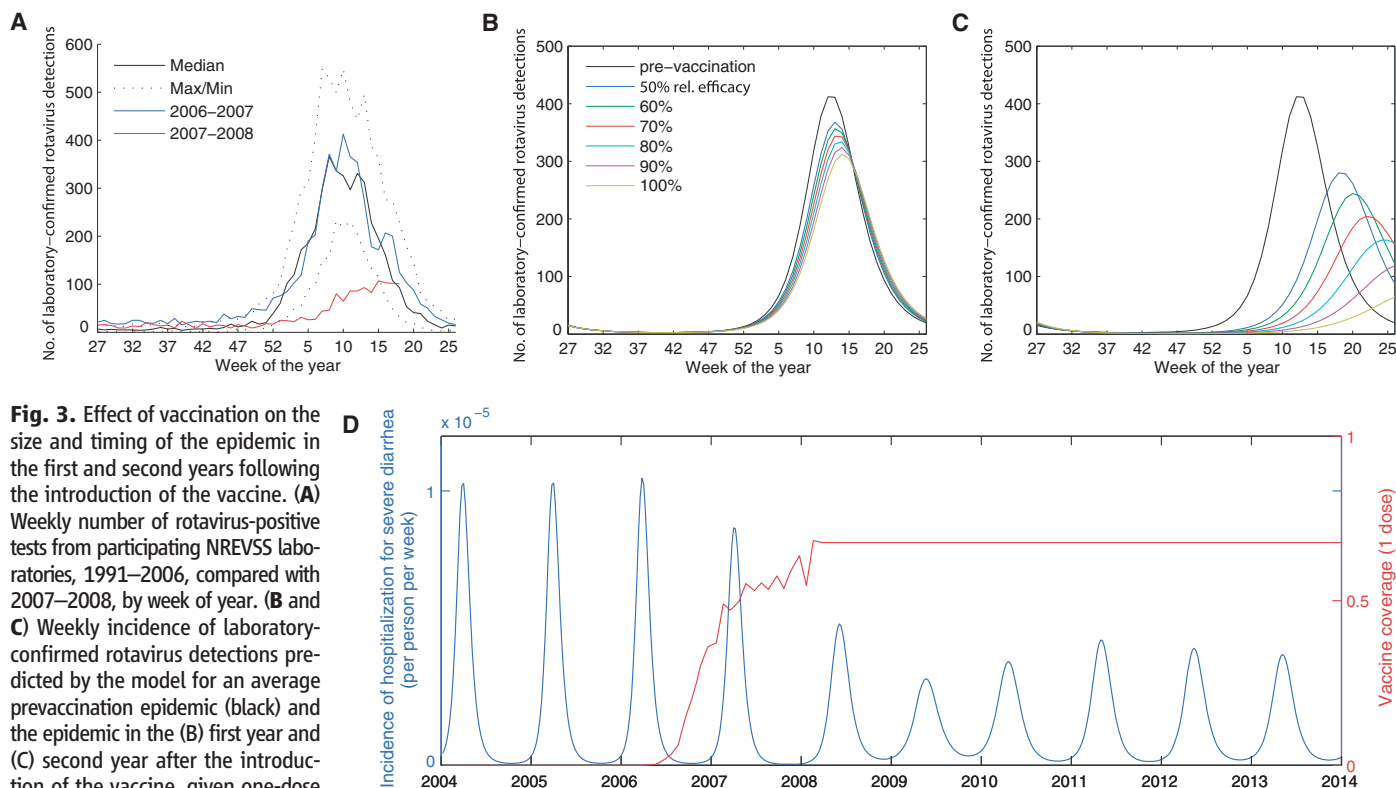


**Fig. 3.** Effect of vaccination on the size and timing of the epidemic in the first and second years following the introduction of the vaccine. (**A**) Weekly number of rotavirus-positive tests from participating NREVSS laboratories, 1991–2006, compared with 2007–2008, by week of year. (**B** and **C**) Weekly incidence of laboratory-confirmed rotavirus detections predicted by the model for an average prevaccination epidemic (black) and the epidemic in the (B) first year and (C) second year after the introduction of the vaccine, given one-dose vaccine coverage estimates and assuming 50 to 100% relative effectiveness. (**D**) Time series of model predictions for effect of vaccination (introduced in 2006) on the incidence of laboratory-confirmed rotavirus (blue), assuming vaccine coverage (red) remains at its current level (~68% with one dose, with 70% relative effectiveness).

infection is not lifelong (28). Our model also indicated that there is little or no indirect protection against severe diarrhea (Fig. 4A). As the age at immunization increases [i.e., if protection only occurs after two or three doses of the vaccine (10)], more individuals can be infected before being fully immunized, thereby reducing the benefit of vaccination. However, the decrease in clinical severity (and thereby the infectiousness) of subsequent infections in vaccinated individuals leads to some indirect protection against infection in general. Under all vaccination schemes, the reduction in the prevalence of infection is greater than that predicted by the direct effect of vaccination alone (Fig. 4B); for example, herd immunity provides an additional 15 to 25% reduction in infection prevalence at 70% vaccine coverage.

The average age of severe diarrhea cases is predicted to increase with higher vaccine coverage (Fig. 4C). The decrease in prevalence of infection in the population delays the time to primary (and secondary) infection such that the average age of severe diarrhea cases increases from 1.5 years old to 4.6 to 7.7 years old, depending on the age at immunization. This increase in the average age of first infection could lead to further decreases in the incidence of severe diarrhea (beyond those predicted by the model) if cases in older children tend to be less severe compared with those in infants (29).

As vaccine coverage level increases, the epidemic peaks occur later than prevaccination, but there tends to be less seasonal variation in the number of rotavirus cases; as a result, the mean timing of rotavirus activity may occur slightly earlier (Fig. 4D). This effect is similar to the birth rate dropping below currently observed levels. As vaccine coverage reaches about 70 to 80%, our model indicates that rotavirus activity follows a pattern of biennial epidemics (reflecting the nonlinear feedback of herd immunity). Still higher vaccine coverage levels can lead to potentially irregular dynamics (particularly if we consider stochastic variations) and a period of years in which there are very few cases of severe diarrhea.
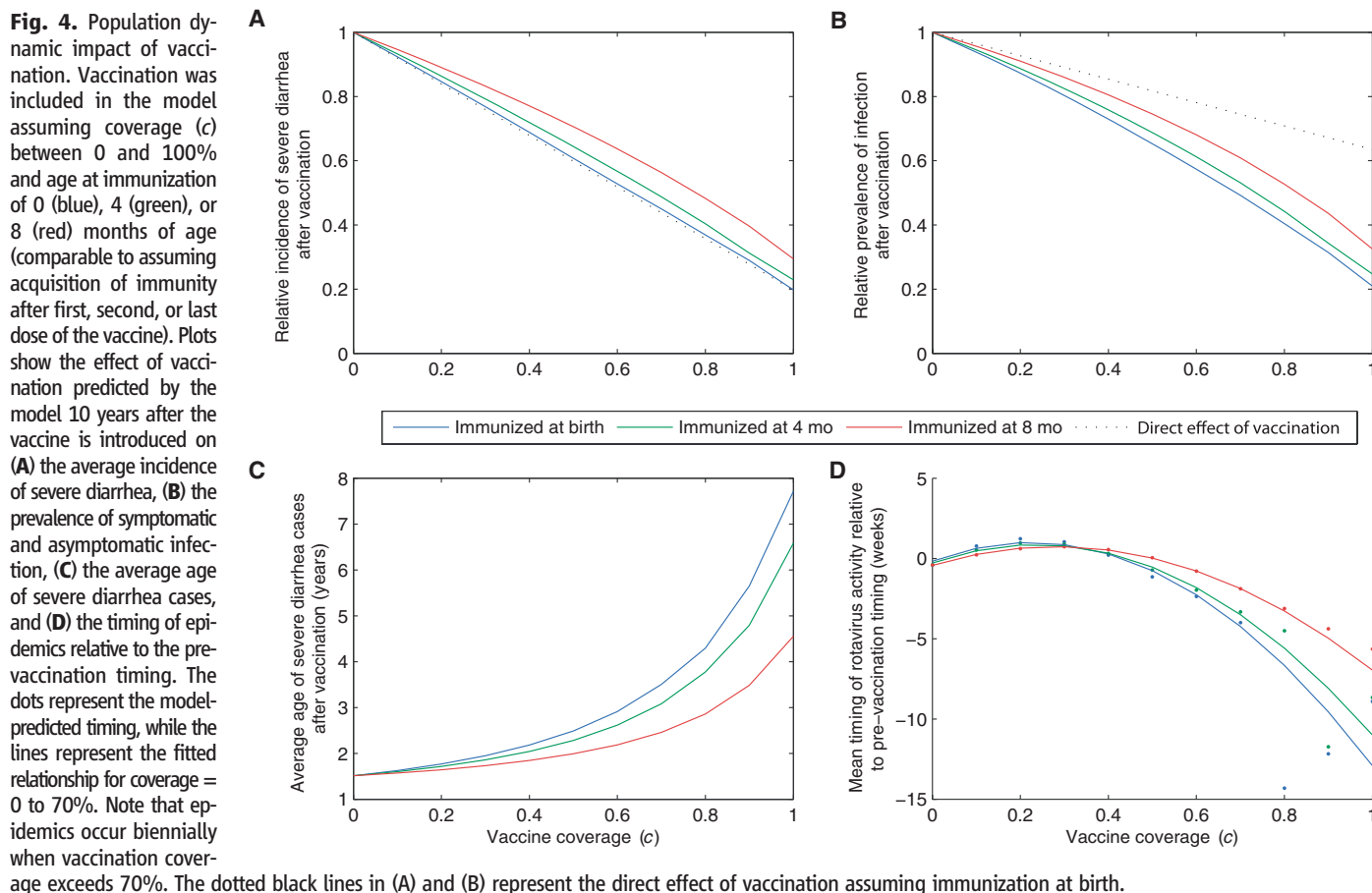
**Discussion.** The analysis of spatiotemporal epidemic patterns has considerable power to help illuminate epidemic dynamics. Traveling waves of infection originating in large population centers have been noted and described for measles (21), rabies (30), dengue (31), and influenza (14). Unlike these pathogens, however, rotavirus infection appears to persist throughout the year even in relatively small populations, possibly through recurring infections in adults (10, 17, 18).

We propose that spatiotemporal variation in birth rate can lead to secular changes in the pattern of rotavirus epidemics. If individuals experiencing their first infection are the primary drivers of epidemics, then demographic changes will have a strong influence on epidemic dynamics. At current rates of susceptible recruitment, small (~5%) seasonal variation in transmissibility can lead to strongly seasonal outbreaks of rotavirus resulting from the annual depletion of infants susceptible to primary infection. The observed variation in birth rates over time and space dictates the relative timing of epidemics across the United States, as evidenced by a state-specific rotavirus transmission model calibrated against rich epidemiological data sets. Qualitatively similar relationships between rotavirus dynamics and the recruitment of susceptibles have been observed in other countries and time periods (10). By contrast, other nondemographic factors may explain why transmissibility tends to be slightly greater in the winter in the United States and elsewhere (10).

The large-scale introduction of pediatric rotavirus vaccination in the United States provides a natural experiment to validate our model, by reducing the recruitment rate of fully susceptible individuals. Incorporating this effect into the model predicts the changes in spatiotemporal dynamics of epidemics associated with early vaccination efforts, verifying our demographic hypothesis. Vaccination against rotavirus has reduced the incidence of severe diarrhea and could lead to changes in the average age of cases and the timing of epidemics. Furthermore, the vaccination program is predicted to decrease the overall prevalence of infection by a degree that is greater than that predicted by the direct effect of vaccination alone.

Although our model reproduced several features of the spatiotemporal dynamics of rotavirus in the United States, further refinement of the model may be necessary (10). We have only implicitly accounted for the interaction between different strains of rota-



**Fig. 4.** Population dynamic impact of vaccination. Vaccination was included in the model assuming coverage (*c*) between 0 and 100% and age at immunization of 0 (blue), 4 (green), or 8 (red) months of age (comparable to assuming acquisition of immunity after first, second, or last dose of the vaccine). Plots show the effect of vaccination predicted by the model 10 years after the vaccine is introduced on (**A**) the average incidence of severe diarrhea, (**B**) the prevalence of symptomatic and asymptomatic infection, (**C**) the average age of severe diarrhea cases, and (**D**) the timing of epidemics relative to the prevaccination timing. The dots represent the model-predicted timing, while the lines represent the fitted relationship for coverage = 0 to 70%. Note that epidemics occur biennially when vaccination coverage exceeds 70%. The dotted black lines in (**A**) and (**B**) represent the direct effect of vaccination assuming immunization at birth.

virus, for which there may be varying degrees of homotypic and heterotypic immunity (32). Secondary infections may occur only when individuals encounter a strain that substantially differs from the one causing primary infection. Furthermore, the true effect of vaccination may differ slightly from that suggested by our model. If vaccination conferred highly protective immunity comparable to that exhibited after two natural infections, our model suggests that the level of herd immunity generated by vaccination could lead to the elimination of the infection from the population at very high coverage levels (10). However, one cannot rule out the possible emergence of new rotavirus strains in response to vaccine pressure, and information on rotavirus genetic diversity will be crucial to understand the long-term effectiveness of any immunization program.

We can extend our U.S.-based analysis to the context of developing countries, where rotavirus remains a substantial cause of childhood morbidity and mortality and disease dynamics differ. The high birth rates typical of developing countries may help explain why rotavirus exhibits less seasonal variation in such settings (33), although climatic factors could also play a role. In addition, rotavirus vaccine efficacy remains somewhat unclear in developing country settings and could be lower than in the United States because of several factors that might interfere with vaccine performance (e.g., presence of maternal antibodies, high levels of coinfection with other enteropathogens, higher rates of malnutrition, and greater prevalence of uncommon rotavirus strains). Efficacy trials of rotavirus vaccines are ongoing in several countries of Asia and Africa, and results are expected in the next 6 to 12 months. Differences in population demographics,

epidemiology of rotavirus disease, and, potentially, vaccine effectiveness, would need to be carefully considered when predicting the benefits of vaccination in developing countries, and the vaccine experience of industrialized nations may not directly translate to countries with high rotavirus mortality burden. Introducing vaccination would likely decrease the overall burden of disease but could have important dynamic consequences, which are key to explore in future research.

**References and Notes**
1. A. Z. Kapikian, R. M. Chanock, in *Fields Virology*, B. N. Fields *et al.*, Eds. (Lippincott-Raven, Philadelphia, 2001), vol. 2, pp. 1657–1708.
2. U. D. Parashar, C. J. Gibson, J. S. Bresse, R. I. Glass, *Emerg. Infect. Dis.* **12**, 304 (2006).
3. U. D. Parashar, E. G. Hummelman, J. S. Bresee, M. A. Miller, R. I. Glass, *Emerg. Infect. Dis.* **9**, 565 (2003).
4. T. K. Fischer *et al.*, *J. Infect. Dis.* **195**, 1117 (2007).
5. R. I. Glass, U. D. Parashar, *N. Engl. J. Med.* **354**, 75 (2006).
6. R. I. Glass *et al.*, *Lancet* **368**, 323 (2006).
7. G. M. Ruiz-Palacios *et al.*, *N. Engl. J. Med.* **354**, 11 (2006).
8. T. Vesikari *et al.*, *N. Engl. J. Med.* **354**, 23 (2006).
9. A. C. Linhares *et al.*, *Lancet* **371**, 1181 (2008).
10. Materials and methods are available as supporting material on *Science* Online.
11. C. W. LeBaron, J. Lew, R. I. Glass, J. M. Weber, G. M. Ruiz-Palacios, *JAMA* **264**, 983 (1990).
12. T. J. Torok *et al.*, *Pediatr. Infect. Dis. J.* **16**, 941 (1997).
13. R. M. Turcios *et al.*, *Pediatr. Infect. Dis. J.* **25**, 451 (2006).
14. C. Viboud *et al.*, *Science* **312**, 447 (2006); published online 29 March 2006 (10.1126/science.1125237).
15. Y. Xia, O. N. Bjornstad, B. T. Grenfell, *Am. Nat.* **164**, 267 (2004).
16. M. I. Nelson, L. Simonsen, C. Viboud, M. A. Miller, E. C. Holmes, *PLoS Pathog.* **3**, 1220 (2007).
17. M. J. Cox, G. F. Medley, *Epidemiol. Infect.* **131**, 719 (2003).
18. H. Nakajima *et al.*, *Lancet* **357**, 1950 (2001).
19. S. Arista *et al.*, *J. Virol.* **80**, 10724 (2006).
20. S. G. Purohit, S. D. Kelkar, V. Simha, *J. Diarrhoeal Dis. Res.* **16**, 74 (1998).
21. B. T. Grenfell, O. N. Bjornstad, J. Kappey, *Nature* **414**, 716 (2001).
22. A. J. Conlan, B. T. Grenfell, *Proc. R. Soc. London Ser. B* **274**, 1133 (2007).
23. G. Cilla, E. Perez-Trallero, M. C. Lopez-Lopategui, A. Gilsetas, M. Gomariz, *Epidemiol. Infect.* **125**, 677 (2000).
24. B. E. Hamilton, P. D. Sutton, S. J. Ventura, "Revised birth and fertility rates for the 1990s and new rates for Hispanic populations, 2000 and 2001: United States" (National Center for Health Statistics, 2003).
25. Centers for Disease Control and Prevention National Center for Health Statistics, VitalStats, www.cdc.gov/nchs/vitalstats/VitalStatsbirths.htm [accessed 22 April 2008].
26. Agency for Healthcare Quality and Research (AHRQ), www.hcup-us.ahrq.gov/databases.jsp.
27. M. A. Staat *et al.*, *MMWR Morb. Mortal. Wkly. Rep.* **57**, 697 (2008).
28. R. M. Anderson, R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford Univ. Press, Oxford, 1991).
29. D. I. Bernstein, D. S. Sander, V. E. Smith, G. M. Schiff, R. L. Ward, *J. Infect. Dis.* **164**, 277 (1991).
30. D. L. Smith, B. Lucey, L. A. Waller, J. E. Childs, L. A. Real, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3668 (2002).
31. D. A. Cummings *et al.*, *Nature* **427**, 344 (2004).
32. B. Jiang, J. R. Gentsch, R. I. Glass, *Clin. Infect. Dis.* **34**, 1351 (2002).
33. S. M. Cook, R. I. Glass, C. W. LeBaron, M. S. Ho, *Bull. World Health Organ.* **68**, 171 (1990).
34. V.E.P and B.G. were supported by NIH (grant R01 GM083983-01) and the Bill and Melinda Gates Foundation. V.E.P, B.G., and L.S. were also supported by the RAPIDD program of the Science and Technology Directorate, U.S. Department of Homeland Security, and the Fogarty International Center, NIH. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention (CDC).

# Nonlocal Transport in the Quantum Spin Hall State

Andreas Roth,[1] Christoph Brüne,[1] Hartmut Buhmann,[1] Laurens W. Molenkamp,[1]* Joseph Maciejko,[2,3] Xiao-Liang Qi,[2,3] Shou-Cheng Zhang[2,3]

Nonlocal transport through edge channels holds great promise for low-power information processing. However, edge channels have so far only been demonstrated to occur in the quantum Hall regime, at high magnetic fields. We found that mercury telluride quantum wells in the quantum spin Hall regime exhibit nonlocal edge channel transport at zero external magnetic field. The data confirm that the quantum transport through the (helical) edge channels is dissipationless and that the contacts lead to equilibration between the counterpropagating spin states at the edge. The experimental data agree quantitatively with the theory of the quantum spin Hall effect. The edge channel transport paves the way for a new generation of spintronic devices for low-power information processing.

The search for topological states of quantum matter has become an important goal in condensed matter physics. Inside a topological insulator, the conventional laws of electrodynamics are substantially altered (1), which may have applications in constructing novel devices for the processing of (quantum) information. The

quantum spin Hall (QSH) state (2, 3) is a topologically nontrivial state of matter that exists in the absence of any external magnetic field. It has a bulk energy gap but gapless helical edge states protected by time reversal symmetry. In the QSH regime, opposite spin states forming a Kramers doublet counterpropagate at the edge (4, 5). Recently, the QSH state was theoretically predicted in HgTe quantum wells (6). There is a topological quantum phase transition at a critical thickness $d_c$ of the quantum well, separating the trivial insulator state for $d < d_c$ from the QSH insulator state for $d > d_c$. Soon after the theoretical prediction, evidence for the QSH state was observed in transport measurements (7). In the QSH regime, experiments measured a conductance $G$ close to twice the quantum unit of conductance $G = 2e^2/h$ (where $e$ is the charge on the electron and $h$ is Planck's constant); this value is consistent with quantum transport due to helical edge states. However, such a conductance quantization in small Hall bar geometries does not allow us to distinguish experimentally between ballistic and edge channel transport in a convincing manner. Thus, it is important to be able to prove experi-

[1]Physikalisches Institut (EP3) and Röntgen Center for Complex Material Systems, Universität Würzburg, Am Hubland, 97074 Würzburg, Germany. [2]Department of Physics, Stanford University, Stanford, CA 94305, USA. [3]Stanford Institute for Materials and Energy Sciences, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA.

*To whom correspondence should be addressed. E-mail: molenkmp@physik.uni-wuerzburg.de

mentally in an unambiguous manner the existence of edge channels in HgTe quantum wells.

**Ohm's law versus nonlocal transport.** In conventional diffusive electronics, bulk transport satisfies Ohm's law. The resistance is proportional to the length and inversely proportional to the cross-sectional area, implying the existence of a local resistivity or conductivity tensor. However, the existence of edge states necessarily leads to nonlocal transport, which invalidates the concept of local resistivity. Such nonlocal transport has been experimentally observed in the quantum Hall (QH) regime in the presence of a large magnetic field (*8*), and nonlocal transport is well described by a quantum transport theory based on the Landauer-Büttiker formalism (*9*). These measurements constitute definitive experimental evidence for the existence of edge states in the QH regime.

We report nonlocal transport measurements in HgTe quantum wells that demonstrate the existence of the predicted extended edge channels. We have fabricated structures more complicated than a standard Hall bar that allow a detailed investigation of the transport mechanism. In addition, we present the theory of quantum transport in the QSH regime, and uncover the effects of macroscopic time irreversibility on the helical edge states.

**Device structure.** We present experimental results on four different devices. The behavior in these structures is exemplary for the ~50 devices we studied. The devices were fabricated from HgTe/(Hg,Cd)Te quantum well structures with well thicknesses of $d = 7.5$ nm (samples D1, D2, and D3) and 9.0 nm (sample D4). Note that all wells have a thickness $d > d_c \approx 6.3$ nm, and thus exhibit the topologically nontrivial inverted band structure. At zero gate voltage, the samples are n-type and have a carrier density of $n_s \approx 3 \times 10^{11}$ cm$^{-2}$ and a mobility of $1.5 \times 10^5$ cm$^2$ V$^{-1}$ s$^{-1}$, with small variations between the different wafers. The devices are lithographically patterned using electron-beam lithography and subsequent Ar ion-beam etching. Devices D1 and D2 are micrometer-scale Hall bars with exact dimensions as indicated in the insets of Fig. 1. Devices D3 and D4 are dedicated structures for identifying nonlocal transport, with schematic structure given in Fig. 2. All devices are fitted with a 110-nm-thick Si$_3$N$_4$/SiO$_2$ multilayer gate insulator and a Ti (5 nm)–Au (50 nm) gate electrode stack.

By applying a voltage $V_g$ to the top gate, the electron carrier density of the quantum well can be adjusted, going from an n-type behavior at positive gate voltages through the bulk insulator state into a p-type regime at negative gate voltages. For reasons of comparison, the experimental data in Figs. 1, 3, and 4 are plotted as a function of a normalized gate voltage $V^* = V_g - V_{thr}$ ($V_{thr}$ is defined as the voltage for which the resistance
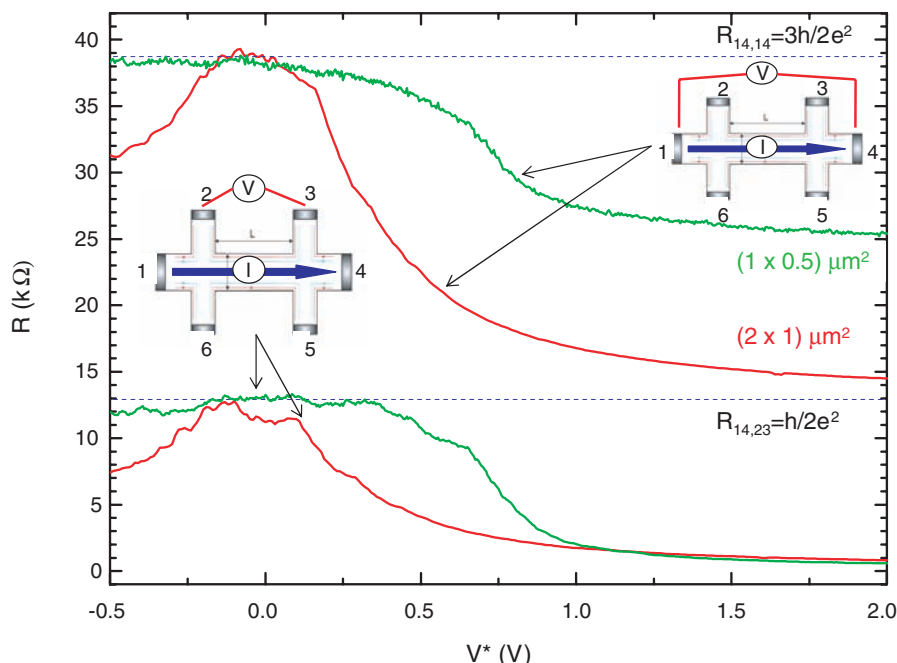


**Fig. 1.** Two-terminal ($R_{14,14}$) (top two traces) and four-terminal ($R_{14,23}$) (bottom traces) resistance versus (normalized) gate voltage for the Hall bar devices D1 and D2 with dimensions (length × width) as indicated. The dotted blue lines indicate the resistance values expected from the Landauer-Büttiker approach.

**Fig. 2.** Schematic layout of devices D3 (**A**) and D4 (**B**). The gray areas are the mesas, the yellow areas the gates, with dimensions as indicated. The numbers indicate the coding of the leads.
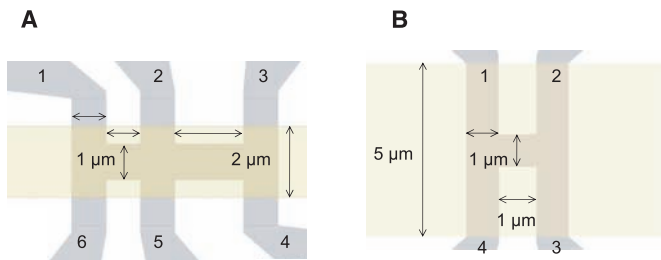


**Fig. 3.** Four- and two-terminal resistance measured on device D3: (**A**) $R_{14,23}$ (red line) and $R_{14,14}$ (green line) and (**B**) $R_{13,54}$ (red line) and $R_{13,13}$ (green line). The dotted blue lines indicate the expected resistance value from a Landauer-Büttiker calculation.
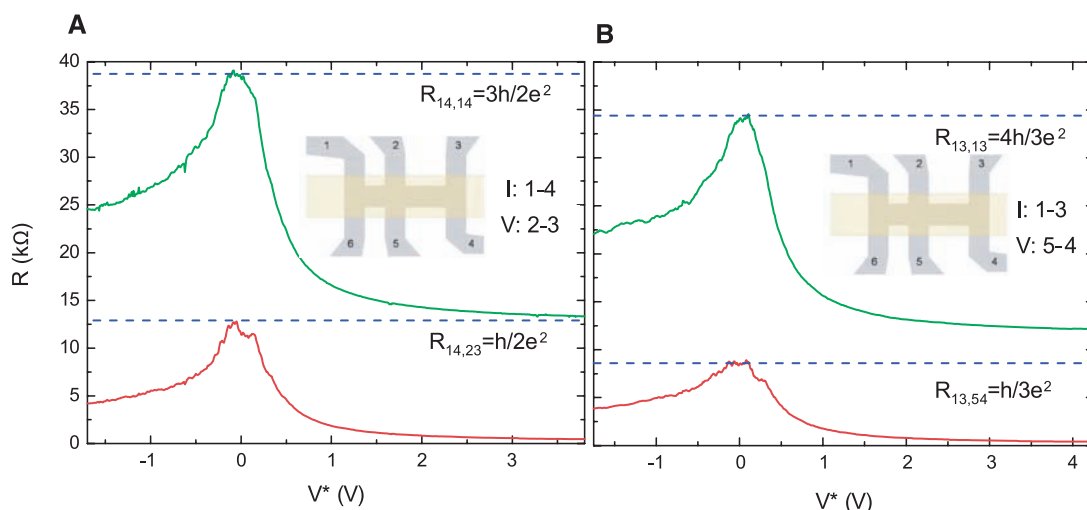
is largest). Measurements were performed at a lattice temperature of 10 mK for samples D1, D2, and D3 and at 1.8 K for sample D4, using low-frequency (13 Hz) lock-in techniques under voltage bias. The four-terminal resistance (Fig. 1) shows a maximum at about $h/2e^2$, in agreement with the results of (7). The contact resistance should be insensitive to the gate voltage and can be measured from the resistance deep in the metallic region. By subtracting the contact resistance, we find that the two-terminal resistance has its maximum at about $3h/2e^2$ (Fig. 1). This value is exactly what is expected from the theory of QSH edge transport obtained from the Landauer-Büttiker formula.

**Transport on the edge.** Within the general Landauer-Büttiker formalism (10), the current-voltage relationship is expressed as

$$I_i = \frac{e^2}{h}\sum_j (T_{ji}V_i - T_{ij}V_j) \qquad (1)$$

where $I_i$ is the current flowing out of the $i$th electrode into the sample region, $V_i$ is the voltage on the $i$th electrode, and $T_{ji}$ is the transmission probability from the $i$th to the $j$th electrode. The total current is conserved in the sense that $\sum_i I_i = 0$. A voltage lead $j$ is defined by the condition that it draws no net current (i.e., $I_j = 0$). The physical currents are left invariant if the voltages on all electrodes are shifted by a constant amount μ, implying that $\sum_i T_{ij} = \sum_i T_{ji}$. In a time reversal–invariant system, the transmission coefficients satisfy the condition $T_{ij} = T_{ji}$.

For a general two-dimensional sample, the number of transmission channels scales with the width of the sample, so that the transmission matrix $T_{ij}$ is complicated and nonuniversal. However, a tremendous simplification arises if the quantum transport is entirely dominated by the edge states. In the QH regime, chiral edge states are responsible for the transport. For a standard Hall bar with $N$ current and voltage leads attached (compare the insets of Fig. 1 with $N = 6$), the transmission matrix elements for the ν = 1 QH state are given by $T(\text{QH})_{i+1,i} = 1$, for $i = 1, …, N$, and all other matrix elements vanish identically. Here we periodically identify the $i = N + 1$ electrode with $i = 1$. Chiral edge states are protected from backscattering; therefore, the $i$th electrode transmits perfectly to the neighboring $(i + 1)$th electrode on one side only. In the example of current leads on electrodes

1 and 4, and voltage leads on electrodes 2, 3, 5, and 6, one finds that $I_1 = -I_4 \equiv I_{14}$, $V_2 - V_3 = 0$, and $V_1 - V_4 = (h/e^2)I_{14}$, giving a four-terminal resistance of $R_{14,23} = 0$ and a two-terminal resistance of $R_{14,14} = h/e^2$.

In the case of helical edge states in the QSH regime, opposite spin states form a Kramers pair, counterpropagating on the same edge. The helical edge states are protected from backscattering due to time reversal symmetry, and the transmission from one electrode to the next is perfect. From this point of view, the helical edge states can be viewed as two copies of chiral edge states related by time reversal symmetry. Therefore, the transmission matrix is given by $T(\text{QSH}) = T(\text{QH}) + T^\dagger(\text{QH})$, implying that the only nonvanishing matrix elements are given by

$$T(\text{QSH})_{i+1,i} = T(\text{QSH})_{i,i+1} = 1 \qquad (2)$$

Considering again the example of current leads on electrodes 1 and 4 and voltage leads on electrodes 2, 3, 5, and 6, one finds that $I_1 = -I_4 \equiv I_{14}$, $V_2 - V_3 = (h/2e^2)I_{14}$, and $V_1 - V_4 = (3h/e^2)I_{14}$, giving a four-terminal resistance of $R_{14,23} = h/2e^2$ and a two-terminal resistance of $R_{14,14} = 3h/2e^2$. The experimental data in Fig. 1 confirm this picture. For both micro–Hall bar structures D1 and D2 that differ only in the dimensions of the area between voltage contacts 3 and 4, we observe exactly the expected resistance values for $R_{14,23} = h/2e^2$ and $R_{14,14} = 3h/2e^2$ for gate voltages where the samples are in the QSH regime.

**Dissipationless transport.** Conceptually, one might sense a paradox between the dissipationless nature of the QSH edge states and the finite four-terminal longitudinal resistance $R_{14,23}$, which vanishes for the QH state. We can generally assume that the microscopic Hamiltonian governing the voltage leads is invariant under time reversal symmetry; therefore, one would naturally ask how such leads could cause the dissipation of the helical edge states, which are protected by time reversal symmetry. In nature, the time reversal symmetry can be broken in two ways: at the level of the microscopic Hamiltonian, or at the level of the macroscopic irreversibility in systems whose microscopic Hamiltonian respects the time reversal symmetry. When the helical edge

states propagate without dissipation inside the QSH insulator between the electrodes, neither form of time reversal symmetry breaking is present. As a result, the two counterpropagating channels can be maintained at two different quasi–chemical potentials, leading to a net current flow. However, once they enter the voltage leads, they interact with a reservoir containing infinitely many low-energy degrees of freedom, and the time reversal symmetry is effectively broken by the macroscopic irreversibility. As a result, the two counterpropagating channels equilibrate at the same chemical potential, determined by the voltage of the lead. Dissipation occurs with the equilibration process. The transport Eq. 1 breaks the macroscopic time reversal symmetry, even though the microscopic time reversal symmetry is ensured by the relationship $T_{ij} = T_{ji}$. In contrast to the case of QH state, the absence of dissipation of the QSH helical edge states is protected by Kramers' theorem, which relies on the quantum phase coherence of wave functions. Thus, dissipation can occur once the phase coherence is destroyed in the metallic leads. By contrast, the robustness of QH chiral edge states does not require phase coherence. The result of a more rigorous and microscopic analysis on the different role played by a metallic lead in QH and QSH states (11) agrees with the simple transport Eqs. 1 and 2. These two equations, which correctly describe the dissipationless quantum transport inside the QSH insulator and the dissipation inside the electrodes, can be subjected to more stringent experimental tests than the two- and four-terminal experiments of Fig. 1 by considering devices D3 and D4 (Fig. 2).

**Helical versus chiral.** A further difference between helical and chiral edge channels is evident from our experiments on the six-terminal device D3 (Fig. 3). When the longitudinal resistance of device D3 is measured by passing a current through contacts 1 and 4 and by detecting the voltage between contacts 2 and 3 ($R_{14,23}$), we find, similar to the results of Fig. 1, a resistance value of $h/2e^2$ when the bulk of the device is gated into the insulating regime (Fig. 3A). However, the longitudinal resistance is markedly different in a slightly modified configuration, where the current is passed through contacts 1 and 3 and the voltage is measured between contacts 5 and 4 ($R_{13,54}$) (Fig. 3B). We now find $R_{13,54} \approx 8.6$ kilohms, which is markedly different from what one would expect for either the QH transport or the purely diffusive transport, where this configuration would be equivalent to the previous one. Application of Eqs. 1 and 2 actually predicts that the observed behavior is indeed what one expects for helical edge channels. This resistance value can again be expressed as an integer fraction of the inverse conductance quanta $e^2/h$: $R_{13,54} = h/3e^2$. This result shows that the current through the device is influenced by the number of ohmic contacts in the current path. These ohmic contacts lead to the equilibration of the chemical potentials between the two counterpropagating helical edge channels inside the contact. There are also some devices for which

**Fig. 4.** Nonlocal four-terminal resistance and two-terminal resistance measured on the H-bar device D4: $R_{14,23}$ (red line) and $R_{14,14}$ (green line). Again, the dotted blue line represents the theoretically expected resistance value.

$R_{14,14} = 3h/4e^2$

$R_{14,23} = h/4e^2$

I: 1-4
V: 2-3

R (kΩ)

V* (V)
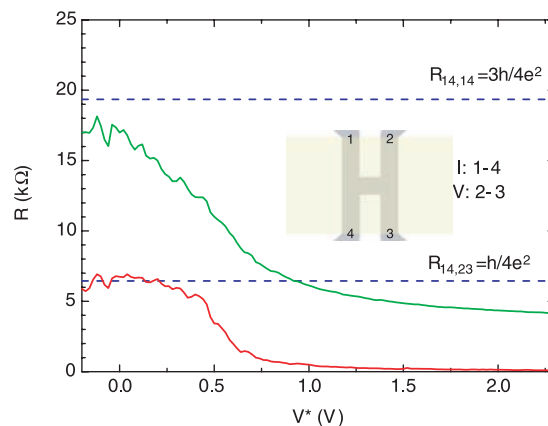
the maximal resistance does not match the theoretical value obtained from Eqs. 1 and 2, but still remains an integer fraction of the quantum $h/e^2$. This result can be naturally understood as due to inhomogeneities in the gate action (e.g., due to interface trap states) inducing some metallic droplets close to the edge channels while the bulk of the sample is insulating. A metallic droplet can cause dephasing of the electronic wave function, leading to fluctuations in the device resistance. For full dephasing, the droplet plays the role of an additional ohmic contact, just as for the chiral edge channels in the QH regime (8). More details on the effects of additional ohmic contacts in the QSH state are given in (11).

Another measurement that directly confirms the nonlocal character of the helical edge channel transport in the QSH regime is in Fig. 4, which shows data obtained from device D4, in the shape of the letter H. In this four-terminal device, the current is passed through contacts 1 and 4 and the voltage is measured between contacts 2 and 3. In the metallic n-type regime (low gate voltage), the voltage signal tends to zero. In the insulating regime, however, the nonlocal resistance signal increases to ~6.5 kilohms, which again fits perfectly to the result of Laudauer-Büttiker considerations: $R_{14,23} = h/4e^2 \approx 6.45$ kilohms. Classically, one would expect only a minimal signal in this configuration (from Poisson's equation, assuming diffusive transport, one estimates a signal of about 40 ohms), and certainly not one that increases so strongly when the bulk of the sample is depleted. This signal measured here is fully nonlocal and can be taken (as was done 20 years ago for the QH regime) as definite evidence of the existence of edge channel transport in the QSH regime. A similar nonlocal voltage has been studied in a metallic spin

Hall system with the same H-bar geometry (12), in which case the nonlocal voltage can be understood as a combination of the spin Hall effect and the inverse spin Hall effect (13). The quantized nonlocal resistance $h/4e^2$ we find here is the quantum counterpart of the metallic case. For example, if we assume that the chemical potential in contact 1 is higher than that in contact 4 (compare to the layout of D4 in Fig. 2B), more electrons will be injected into the upper edge state in the horizontal segment of the H-bar than into the lower edge state. Because on opposite edges the right-propagating edge states have opposite spin, this implies that a spin-polarized current is generated by an applied bias $V_1 - V_4$, comparable to a spin Hall effect. When this spin-polarized current is injected into the right leg of the device, the inverse effect occurs. Electrons in the upper edge flow to contact 2 while those in the lower edge flow to contact 3, establishing a voltage difference between those two contacts due to the charge imbalance between the edges. The right leg of the device thus acts as a detector for the injected spin-polarized current, which corresponds to the inverse spin Hall effect.

**Concluding remarks.** The multiterminal and nonlocal transport experiments on HgTe microstructures in the QSH regime demonstrate that charge transport occurs through extended helical edge channels. We have extended the Landauer-Büttiker model for multiterminal transport in the QH regime to the case of helical QSH edge channels and have shown that this model convincingly explains the observations. Logic devices based on the complementary metal oxide semiconductor design generate considerable heating due to the ohmic dissipation within the channel. Our work on conductance quantization demonstrates that electrons can be transported coherently

within the edge channel without ohmic dissipation. Such an effect can be used to construct logic devices with improved performance.

**References and Notes**
1. S. C. Zhang, *Physics* **1**, 6 (2008).
2. C. L. Kane, E. J. Mele, *Phys. Rev. Lett.* **95**, 226801 (2005).
3. B. A. Bernevig, S. C. Zhang, *Phys. Rev. Lett.* **96**, 106802 (2006).
4. C. Wu, B. A. Bernevig, S. C. Zhang, *Phys. Rev. Lett.* **96**, 106401 (2006).
5. C. Xu, J. Moore, *Phys. Rev. B* **73**, 045322 (2006).
6. B. A. Bernevig, T. L. Hughes, S.-C. Zhang, *Science* **314**, 1757 (2006).
7. M. König et al., *Science* **318**, 766 (2007); published online 19 September 2007 (10.1126/science.1148047).
8. C. W. J. Beenakker, H. van Houten, *Solid State Phys.* **44**, 1 (1991).
9. M. Büttiker, *Phys. Rev. B* **38**, 9375 (1988).
10. M. Büttiker, *Phys. Rev. Lett.* **57**, 1761 (1986).
11. See supporting material on *Science* Online.
12. C. Brüne et al., http://arxiv.org/abs/0812.3768 (2008).
13. E. M. Hankiewicz, L. W. Molenkamp, T. Jungwirth, J. Sinova, *Phys. Rev. B* **70**, 241301(R) (2004).
14. We thank T. Beringer, N. Eikenberg, M. König, and S. Wiedmann for assistance in some of the experiments. Supported by Deutsche Forschungsgemeinschaft grant SFB 410; German-Israeli Foundation grant I-881-138.7/2005; NSF grant DMR-0342832; the Alexander von Humboldt Foundation; the U.S. Department of Energy, Office of Basic Energy Sciences, under contract DE-AC03-76SF00515; the Focus Center Research Program (FCRP) Center on Functional Engineered Nanoarchitectonics; the National Science and Engineering Research Council of Canada; and the Stanford Graduate Fellowship Program. Computational work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (www.sharcnet.ca).

# Higher-Order Photon Bunching in a Semiconductor Microcavity

M. Aßmann,[1] F. Veit,[1] M. Bayer,[1]* M. van der Poel,[2] J. M. Hvam[2]

Quantum mechanically indistinguishable particles such as photons may show collective behavior. Therefore, an appropriate description of a light field must consider the properties of an assembly of photons instead of independent particles. We have studied multiphoton correlations up to fourth order in the single-mode emission of a semiconductor microcavity in the weak and strong coupling regimes. The counting statistics of single photons were recorded with picosecond time resolution, allowing quantitative measurement of the few-photon bunching inside light pulses. Our results show bunching behavior in the strong coupling case, which vanishes in the weak coupling regime as the cavity starts lasing. In particular, we verify the $n$ factorial prediction for the zero-delay correlation function of $n$ thermal light photons.

The discovery of two-photon bunching in thermal light by Hanbury Brown and Twiss (1) marked a turning point for the development of quantum optics (2) and has also found applications in a variety of fields, from particle physics (3) to ultracold quantum gases (4). Photon bunching is the tendency of indistinguishable photons, emitted by a thermal or chaotic light source, to show

an enhanced joint detection probability compared with statistically independent particles that are emitted, for instance, by lasers. The explanation of this bunching relies on quantum interference between indistinguishable $n$ particle probability amplitudes leading to excess joint detections if the photon number follows the Bose-Einstein distribution (5, 6).

The quantity describing bunching for two photons is the second-order intensity correlation function defined as

$$g^{(2)}(t,\tau) = \frac{\langle :\hat{n}(t)\hat{n}(t+\tau): \rangle}{\langle \hat{n}(t)\rangle\langle \hat{n}(t+\tau)\rangle} \qquad (1)$$

where $\hat{n} = \hat{a}^\dagger \hat{a}$ is the photon number operator, $t$ and $t + \tau$ are the detection times of the two

[1]Experimentelle Physik II, Technische Universität Dortmund, D-44221 Dortmund, Germany. [2]DTU Fotonik, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark.

*To whom correspondence should be addressed: E-mail: manfred.bayer@tu-dortmund.de

photons, and the double stops denote normal and apex ordering of the underlying photon creation and annihilation operators $\hat{a}^{\dagger}$ and $\hat{a}$. This normal ordering accounts for the change of the light field by the detection of a photon (i.e., when one photon is destroyed). $g^{(2)}(t,\tau)$ gives the average value of joint detections normalized by the product of the average photon numbers at times $t$ and $t + \tau$. The latter is the number of joint detections one would expect if the photons were statistically independent particles. In particular, the equal time correlation function $g^{(2)}(t,0)$ is a direct quantitative measure of the joint detection probability.

Bunching is not limited to two-particle probability amplitudes, but it can be extended to the interference of $n$ particle probability amplitudes. The second-order correlation function can be extended to arbitrary higher orders, thereby providing the most complete characterization of the light field possible (7). The $n$th-order correlation function is the average value of the joint detections of $n$ photons at times $t_i$ (where $i = 1,...,n$) divided by the average photon numbers at the respective times

$$g^{(n)}(t_1...t_n) = \frac{\left\langle :\prod_{i=1}^{n} \hat{n}(t_i): \right\rangle}{\prod_{i=1}^{n} \langle \hat{n}(t_i) \rangle} \qquad (2)$$

Because of the factorial increase of possible permutations of amplitudes with increasing $n$ for indistinguishable photons, bunching is expected to be even more prominent in higher orders: The increase of the equal time joint detection probability of $n$ indistinguishable photons compared with the case of statistically independent particles is predicted to follow an $n$ factorial dependence (8). In this general form, the definition applies to stationary light fields as well as to pulsed light fields. The only difference is that the average in $g^{(n)}$ denotes a time average in the stationary case, whereas it is an average over an ensemble of equal pulses in the nonstationary case. We will focus on equal time correlations; that is, all $t_i$ in the definition of $g^{(n)}$ are the same, and we average over all times $t$. We will refer to this fixture as $g^{(n)}(\tau = 0)$.

However, photons are only indistinguishable if the delays between their emission times do not exceed the coherence time. For larger delays, photon distinguishability switches the photon number distribution to a Poissonian one, which is the distribution of statistically independent particles. In this case, no bunching will occur, and $g^{(n)}(\tau = 0)$ will equal 1 for all orders of $n$.

Usual schemes for measuring photon bunching need precise alignment of $n$ separate detectors and suffer from insufficient time resolution to measure the real $g^{(n)}(\tau = 0)$ (9), as the coherence time is on the order of picoseconds (10). Additionally, most of these schemes are only suitable for stationary light fields or measure the joint detections without any normalization (11) or consecutive photon pairs (12). In all of these cases, one needs to make assumptions about

$g^{(n)}(\tau = 0)$ to reconstruct it (13). It is therefore desirable to be able to access photon bunching directly on short time scales.

Here we studied $g^{(n)}(\tau = 0)$ of photons from the radiative decay of microcavity polaritons. At low densities, microcavity polaritons can be considered as bosons, which are created by the strong coupling of photons confined in a cavity to excitons confined in a semiconductor quantum well. The resulting composite boson is of a mixed excitonic and photonic nature. Accordingly, the far field emission of the microcavity gives information about the coherence properties of the polaritons as the emitted photons are a part of the polaritonic wave function (14). As a consequence of their steep dispersion, polaritons have a small effective mass leading to a low density of states and also to a high-state occupancy, which is a prerequisite for efficient demonstration of photon bunching of a single mode. Unfortunately, the small effective mass also results in a short lifetime of the polaritons, on the picosecond scale. The coherence time is expected to be on the order of picoseconds, which is much shorter than the time needed to reach thermal equilibrium with the lattice.

Nevertheless, in a regime where the ground state is populated dominantly by direct carrier scattering and polariton-polariton scattering is suppressed, one can still achieve emission with thermal characteristics [supporting online material (SOM) text (15)].

In this regime, one expects the microcavity to be a chaotic light source with $g^{(n)}(\tau = 0) = n!$. Under high excitation power, polaritons no longer follow the model of weakly interacting bosons because of the fermionic nature of electrons and holes. Coulomb interactions play a pronounced role with increasing polariton density and manifest as a substantial blue shift of the lower polariton (LP) branch up to the point where the strong coupling regime is bleached. The microcavity turns into a vertical cavity surface-emitting laser, where population inversion and, thus, conventional photon lasing occurs. In this regime, one expects $g^{(n)}(\tau = 0) = 1$ for all orders of $n$.

We present experimental results of a photon-bunching measurement scheme using a streak camera, which directly measures all orders of $g^{(n)}(\tau = 0)$ within a light pulse with a time resolution sufficient for semiconductor light sources. We use a redesigned streak camera, in which the
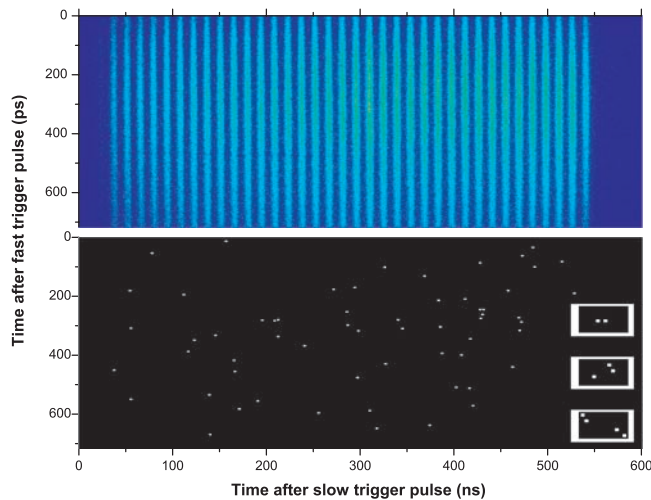


**Fig. 1.** Integrated intensity of 200,000 pictures compared with a single snapshot. **(Insets)** Exemplaric enlarged binning areas of 10 ps containing two-, three-, and four-photon combinations.
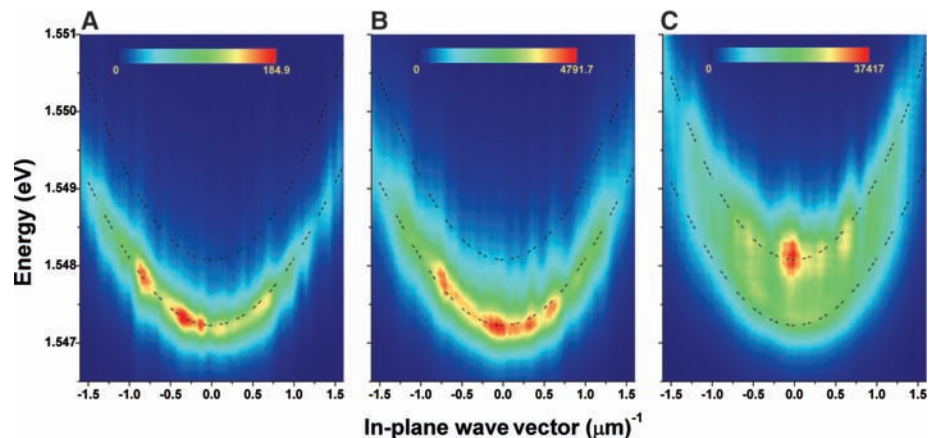


**Fig. 2.** Momentum distribution of the polaritons as measured by angle-resolved photoluminescence for three different excitation densities: **(A)** 50 μW (far below the lasing threshold), **(B)** 1.5 mW (at the lasing threshold), and **(C)** 10 mW (above threshold). The false color scale is linear. The black dashed lines indicate the dispersion of the LP and the bare cavity mode.

photons hit a photocathode and are converted into photoelectrons. These photoelectrons are then accelerated by time-dependent voltages and hit a phosphorus screen. Single pictures of the afterglow of this screen are recorded by a charge-coupled device's (CCD) camera. Because of the precisely controllable ramp voltage, the vertical position of the electron hitting the phosphorus screen is a direct measure of the arrival time of the photon with picosecond time resolution. Another slow, horizontal ramp voltage allows us to record the emission after several pulsed excitations on one screen, with the horizontal position of the photon hitting the phosphorus screen enabling us to distinguish between different pulses. Then pixel binning is performed. The horizontal width of the binning area is determined by the pulse width on the CCD. The vertical length of the binning area determines the time window considered as simultaneous and, therefore, also the

time resolution of the system. In this case, we chose a time window of 10 ps. To determine $g^{(n)}(\tau = 0)$ from the experimental data, we count all $n$-photon combinations in the single pulses and divide this value by the expected number of $n$-photon combinations, which is the product of the averaged intensities at $n$ times. The averaged intensity is easily accessible by summing up all single pictures; we do not need to make any additional assumptions. The results are directly derived from the measured data sets. A single screen compared with the integrated intensity of all screens is shown in Fig. 1. The insets show some binning areas containing two-, three-, and four-photon combinations. See the SOM for further technical details on the experimental setup.

The sample we studied consists of a GaAs/AlGaAs microcavity grown by molecular beam epitaxy. It contains one 10-nm-wide quantum well placed in the electric field antinode of a slightly wedged λ cavity especially designed to

avoid charge accumulation in the quantum well (here, λ is the wavelength of the light confined in the cavity) (16). The sample displays a vacuum Rabi splitting of 3.9 meV. The polariton dispersion for different excitation densities (Fig. 2) shows an apparent bleaching of the strong coupling regime with increasing excitation power. Additionally, we found the LP ground state to be only weakly populated far below the lasing threshold. Therefore, polariton-polariton scattering is also weak in this regime. We investigated the far-field emission of the LP branch at a negative detuning of –2 meV. The Fourier plane of the emission was either imaged onto the entrance slit of a monochromator for measuring the dispersion or onto the entrance slit of a streak camera for photon counting measurements. Photons, which are emitted at an angle of θ, directly correspond to polaritons with energy $E$ and in-plane wave vector of $k_\| = \frac{E}{\hbar c}\sin\theta$ (where $\hbar$ is Planck's constant h divided by $2\pi$ and c is the speed of light). Thus, in the first case, the entrance slit of the monochromator selects a narrow stripe with $k_{x,\|} = 0$. In the second case, only the $k_\| = 0$ state of the LP branch is selected with an angular resolution of ~1° by using a pinhole. Additionally, an interference filter with a 1-nm width is used to ensure that only a single mode contributes to the signal. With increasing excitation density, the filter is tuned so that the central transmission wavelength follows the blue shift of the polariton dispersion as shown in Fig. 3.

The measured time-averaged normalized intensity correlation functions $g^{(n)}(\tau = 0)$ up to the fourth order (Fig. 4) show that, for high excitation densities, all orders approach the expected value of 1, denoting conventional photon lasing. With decreasing excitation density, a smooth transition toward the thermal regime occurs, which is accompanied by photon bunching. At an excitation power of ~1.5 mW, the bunching effect saturates at values of approximately 2 and 6, which are the expected values of $n$ factorial for the second and third orders of $g^{(n)}(\tau = 0)$. The fourth order also shows an increase of the joint detections, but the number of detected four-photon combinations is too small at low excitation densities to give statistically significant results in the thermal light regime. The results for different orders of $g^{(n)}(\tau = 0)$ at the same excitation power are derived from the same data set.

To assure that we measured single-mode thermal emission from the $k_\| = 0$ state, we also increased the collection angle by opening the pinhole. By doing so, we increase the number of modes contributing to the signal and, therefore, leave the regime of indistinguishable photons. As can be seen in the right inset of Fig. 4, photon bunching is only present at collection angles below 1.5°, suggesting that we are indeed operating in the single-mode thermal regime. To further ensure that our experimental results are the result of photon bunching and not just a consequence of some dominating noise source when the signal gets weaker, we also studied the blue shift of the LP and the input-output curve of the microcavity, as



**Fig. 3.** Black squares indicate integrated intensity of the lasing mode measured at normal incidence as a function of the nonresonant excitation power at a detuning of –2 meV. Blue circles represent emission energy around normal incidence as a function of the excitation energy. Dashed lines indicate the linear dependence of the emitted intensity on the excitation power below (lower curve) and above (upper curve) the lasing threshold.



**Fig. 4.** Second- (black triangles), third- (red spheres), and fourth-order (green open circles) intensity correlation function versus excitation power. (**Left inset**) Close-up of the second- and third-order intensity correlation on a normalized linear scale. A value of 0 corresponds to $g^{(n)}(\tau = 0) = 1$, and a value of 1 corresponds to $g^{(n)}(\tau = 0) = n!$. (**Right inset**) Second-order intensity correlation function in the thermal regime for several total collection angles. Error bars indicate the variation of the correlations, resulting from the SDs of the detected photon numbers used for calculating $g^{(2)}$ and $g^{(3)}$.

shown in Fig. 3. The onset of the decrease of $g^{(n)}(\tau = 0)$ coincides with the beginning of the LP blue shift and the onset of a nonlinear increase in the input-output curve. This shows that the system leaves the strong coupling regime and starts to lase. At high excitation powers, well-defined lasing at the bare cavity mode builds up as expected (Fig. 2). It is obvious that the thresholds, where the $g^{(n)}(\tau = 0)$ begin to decrease toward a value of 1, do not occur at the same excitation density. This shift can be explained in terms of the low photon numbers inside the cavity at the lasing threshold. Stimulated emission sets in at a mean photon number $p$ of the order of unity inside the mode of interest, but in the threshold region, there is still a superposition of thermal and stimulated emission present. Because of the stronger photon number fluctuations in chaotic fields, their contribution to $n$-photon combinations will still be substantial, whereas $p$ is smaller than $n$.

Our results verify that, under nonresonant excitation, the ground state of a semiconductor microcavity in the strong coupling regime is a single-mode thermal light source and shows particularly pronounced bunching effects, even in higher orders. In terms of applications, our findings are promising for applications in quantum-optical coherence tomography or ghost imaging optics and allow for thermal light imaging, offering high temporal resolution and massively improved sensitivity for $n$-photon processes as compared with coherent light. From a more fundamental point of view, the demonstrated experimental technique is a versatile tool for studying quantum fluctuations and phase transitions and might especially provide further insight into the intensely debated physics of polariton lasers and the spontaneous phase transition toward polariton Bose-Einstein condensates ([17]).

### References and Notes

1. R. Hanbury Brown, R. Q. Twiss, *Nature* **177**, 27 (1956).
2. F. T. Arecchi, E. Gatti, A. Sona, *Phys. Lett.* **20**, 27 (1966).
3. G. Baym, *Act. Phys. Pol. B* **29**, 1839 (1998).
4. S. Fölling et al., *Nature* **434**, 481 (2005).
5. U. Fano, *Am. J. Phys.* **29**, 539 (1961).
6. G. Scarcelli, V. Berardi, Y. Shih, *Phys. Rev. Lett.* **96**, 063602 (2006).
7. U. M. Titulaer, R. J. Glauber, *Phys. Rev.* **140**, B676 (1965).
8. G. S. Agarwal, *Phys. Rev. A* **1**, 1445 (1970).
9. B. L. Morgan, L. Mandel, *Phys. Rev. Lett.* **16**, 1012 (1966).
10. S. M. Ulrich et al., *Phys. Rev. Lett.* **98**, 043906 (2007).
11. P. Michler et al., *Science* **290**, 2282 (2000).
12. J. Kasprzak et al., *Phys. Rev. Lett.* **100**, 067402 (2008).
13. D. M. McAlister, M. G. Raymer, *Phys. Rev. A* **55**, R1609 (1997).
14. V. Savona, F. Tassone, C. Piermarocchi, A. Quattropani, P. Schwendimann, *Phys. Rev. B* **53**, 13051 (1996).
15. Materials and methods are available as supporting material on *Science* Online.
16. J. R. Jensen, P. Borri, W. Langbein, J. M. Hvam, *Appl. Phys. Lett.* **76**, 3262 (2000).
17. P. Schwendimann, A. Quattropani, *Phys. Rev. B* **77**, 085317 (2008).
18. We thank C. B. Sørensen for growing the wafer and I. Akimov for discussions. This work was supported by the Deutsche Forschungsgemeinschaft research group "Quantum optics in semiconductor nanostructures." A patent application for the optical setup presented here is currently being filed.

# Band Formation from Coupled Quantum Dots Formed by a Nanoporous Network on a Copper Surface

Jorge Lobo-Checa,[1]*†‡ Manfred Matena,[1]† Kathrin Müller,[2] Jan Hugo Dil,[3,4] Fabian Meier,[3,4] Lutz H. Gade,[5] Thomas A. Jung,[2]‡ Meike Stöhr[1]‡

The properties of crystalline solids can to a large extent be derived from the scale and dimensionality of periodic arrays of coupled quantum systems such as atoms and molecules. Periodic quantum confinement in two dimensions has been elusive on surfaces, mainly because of the challenge to produce regular nanopatterned structures that can trap electronic states. We report that the two-dimensional free electron gas of the Cu(111) surface state can be trapped within the pores of an organic nanoporous network, which can be regarded as a regular array of quantum dots. Moreover, a shallow dispersive electronic band structure is formed, which is indicative of electronic coupling between neighboring pore states.

The electronic and optical properties of crystalline solids exhibit properties that derive to a large extent from the periodic arrangement and interactions of their component quantum systems, such as atoms or molecules. Extending the principle of such periodic coupling beyond the molecular regime has given rise to metamaterials, which are composed of regularly repeated units ([1]), in most cases, nanoparticles ([2, 3]). Quantum effects that arise from confinement of electronic states have been extensively studied for surface states of noble metals, which are characterized by a quasi–two-dimensional (2D) electron gas. These may be visualized by scanning tunneling microscopy (STM) as standing wave patterns arising from scattering at steps and defects ([4, 5]) or at large organic molecules ([6]). Examples of such surface state confinement comprise thin films ([7]), artificial nanoscale structures ([8, 9]), vacancy and ad-atom islands ([10, 11]), self-assembled 1D chains ([12, 13]), vicinal surfaces ([14–16]), and quantum dots ([2, 17]).

In spite of these previous examples, periodic quantum confinement in 2D at surfaces has always been elusive, mainly because of the difficulties encountered in the production of strictly regular nanopatterned structures that can trap electronic states. However, we note that Collier et al. observed coupling phenomena between quantum dots at a solid-liquid interface, namely between colloidal particles arranged in a Langmuir monolayer ([17]).

Periodic confinement in 2D is expected to induce regularly distributed discrete energy levels that could be experimentally observed through the appearance of nondispersive subbands, as previously reported for thin films ([7]) or 1D systems ([13, 15, 16]). The size of the confining entities embedded within the 2D periodic nanostructures should be larger than or comparable to a critical length of ~2 nm, as experimentally observed for 1D structures ([15]). The design of such structures is more readily achieved by using molecules as building blocks rather than atomic units, given the fundamental dimensions of these arrays. Potential candidates for molecular systems that might exhibit this zero-dimensional (0D) periodic electronic confinement are porous molecular surface networks. Their production is based on molecular self-assembly, which makes use of concepts established in supramolecular chemistry and has the advantage that identical parts are produced at once. This is in contrast, for instance, to assembly based on atom-by-atom positioning techniques ([8, 9]). Self-assembled nanoporous networks have been obtained by using either hydrogen bonding motifs ([18]) or metal-complexation ([19]) on metal surfaces. Within the pores of these molecular nanoporous networks, electronic confinement is to be expected ([20]).

We report on the interplay of the surface state electrons of Cu(111) with a supramolecular porous network adsorbed on the Cu surface that leads to the formation of a 2D electronic band structure through the coupling of confined elec-

[1]Department of Physics, University of Basel, CH-4056 Basel, Switzerland. [2]Laboratory for Micro- and Nanotechnology, Paul Scherrer Institute, CH-5232 Villigen PSI, Switzerland. [3]Physik-Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. [4]Swiss Light Source, Paul Scherrer Institute, CH-5232 Villigen, Switzerland. [5]Anorganisch-Chemisches Institut, Universität Heidelberg, Im Neuenheimer Feld 270, 69120 Heidelberg, Germany.

*Present address: Centre d'Investigaciò en Nanociència i Nanotecnologia (CIN2), 08193 Bellaterra (Barcelona), Spain.
†These authors contributed equally to this work.
‡To whom correspondence should be addressed. E-mail: jorge.lobo@cin2.es (J.L.-C.); thomas.jung@psi.ch (T.A.J.); meike.stoehr@unibas.ch (M.S.)

tronic states. We chose a perylene derivative, 4,9-diaminoperylene-quinone-3,10-diimine (DPDI), as the organic building block (Fig. 1A), which forms a highly ordered network on Cu(111) upon thermal dehydrogenation (21). By choice of the appropriate molecular coverage before annealing (22) a highly stable hexagonal p(10x10) organic superstructure is generated (Fig. 1B) (21, 23). Each pore of the network confines the surface state of the Cu substrate in what can be described as a 0D quantum dot. Because of the imperfect confinement observed for all 0D cases studied so far on surfaces (8–12), coupling of quantum dots with their neighbors results in shallow dispersive electronic bands. Our results may suggest possible modifications of these artificially created electronic structures, such as changes to the dimensions of the molecular network periodicities

together with the appropriate choice of the substrate.

We used scanning tunneling spectroscopy (STS) to probe the local (electronic) density of states (LDOS) (24). Spectra recorded inside a pore display a peak at –0.22 V that is not seen on the bare metal (Fig. 2A). Given the previous observations of discrete states in 2D nanostructures (8–11), we assign this peak to the confinement of the surface state electrons inside the pore. The confinement of this discrete state is demonstrated by recording simultaneously topography (Fig. 2B) and $dI/dV$ (differential conductance) signal (Fig. 2C) at the peak energy, the latter showing in first order the local distribution of this state. Because each pore features a confined state, it can be considered as a single quantum dot that confines electrons in all three directions. The inherent periodicity of the

molecular network generates a regular "quantum dot array" (Fig. 2C).

The width of the STS peak in Fig. 2A suggests imperfect confinement, which has been previously observed for isolated 2D nanocavities (8–11) as well as for 1D confinement (12). Such cases can be explained by lossy scattering at the confining boundary; inelastic electron-electron and electron-phonon scattering in small nanocavities are thought to be less important (11). Likewise, for our 2D periodic dehydro-DPDI network, which is commensurate with the Cu(111) substrate, the confined electronic states are expected to "leak" into neighboring pores. This leakage would electronically couple quantum wells of neighboring pores, and this coupling should be visible in the electronic structure of the system.

The electronic structure of our quantum dot array was investigated with angle-resolved photoemission spectroscopy (ARPES). This laterally averaging surface-sensitive technique determines the binding energy of the occupied states of the system as a function of the electron momentum. In Fig. 3A, ARPES normal emission spectra recorded at room temperature (RT) are shown for the clean Cu substrate and for the Cu substrate covered by different amounts of the dehydro-DPDI network. For the extreme cases (the clean Cu substrate and the highest molecular coverage), the spectra display a single component line shape, whereas all the other spectra exhibit a double-peak line shape. A detailed analysis of the spectra (fitted by using two Lorentzian components together with a constant background and multiplied by a Fermi-Dirac distribution) shows that the double peak consists of the components observed for the extreme cases, whereas only their relative intensities vary depending on the network coverage (Fig. 3B). The component marked red is attributed to the clean surface state and maintains its binding energy and width
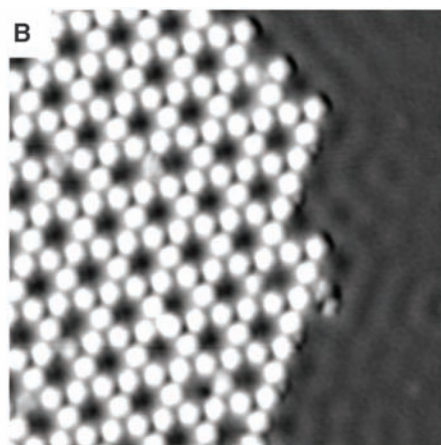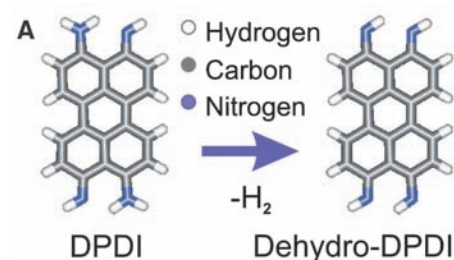


**Fig. 1.** Interaction of the dehydro-DPDI network with the surface state of the underlying substrate. The formation of this porous network is based on a thermally induced dehydrogenation of DPDI on Cu(111). (**A**) Molecular structure of DPDI and its dehydrogenated form. (**B**) STM image for sub-monolayer coverage of DPDI deposited on Cu(111) after thermal annealing at 500 K. Dehydro-DPDI acts as both hydrogen-bond donor and acceptor to produce a very stable self-assembled porous network (image size: 29 nm by 29 nm, –0.1 V, 50 pA) (21). The network periodicity is 2.55 nm with a pore diameter of ~1.6 nm. Standing wave patterns in the Cu surface state arise from the scattering of the delocalized electronic states at the border of the adsorbate adlayer. This can be observed at low tunneling bias (4–6).
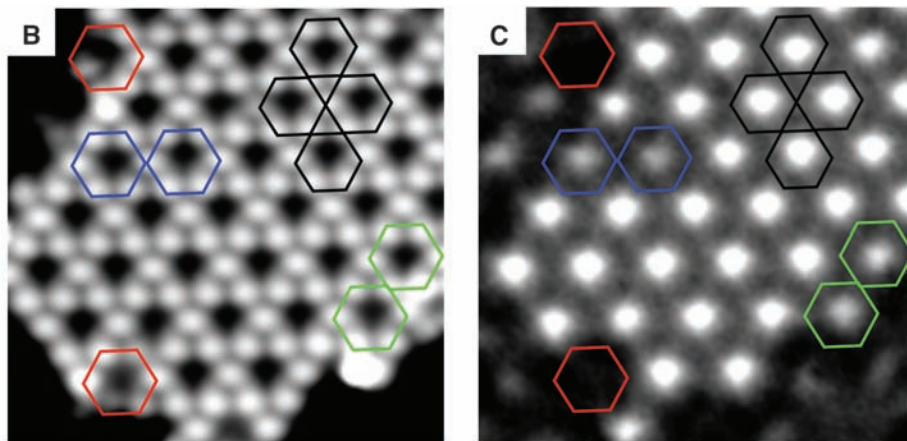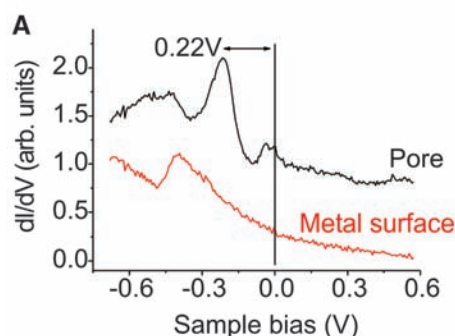


**Fig. 2.** Study of the electronic confinement of the surface state within the pores of the network, corresponding to an array of 0D quantum dots. (**A**) STS spectra obtained at 5 K on the clean Cu surface (red) and inside a pore of the dehydro-DPDI network (black). The latter spectrum exhibits a maximum at –0.22 V, which is attributed to a confined surface state (lock-in: $V_{rms}$ = 6 mV, $f$ = 513 Hz, initial tip parameters: –0.8 V/80 pA). (**B**) STM image of the porous network (13.6 nm by 13.6 nm, –0.20 V, 70 pA) and (**C**) simultaneously recorded $dI/dV$ map at 5 K (lock-in: $V_{rms}$ = 8 mV, $f$ = 513 Hz). A confined electronic state is observed inside the molecular pores. The black

hexagons highlight the molecular positions. Defective pores severely affect the confined state (red hexagons), although not every defect annihilates the confined state (green hexagons). In some cases, the confined state is affected although topography does not indicate any defects (blue hexagons). These observations could be explained by irregularities of the substrate within the pore (e.g., trapped Cu ad-atoms or holes).

**Fig. 3.** Band dispersion resulting from the periodic influence of the porous network on the surface state studied with ARPES for different molecular coverages. (**A**) Normal emission spectra at RT showing the emerging confined state, which progressively replaces the original Cu(111) surface state (scaled, black dotted line) with increasing molecular coverage. (**B**) Peak area, binding energy (B.E.), and full width at half maximum (FWHM) resulting from the analysis of the pristine and the confined states presented in (A). (**C**) Energy dispersion curves (EDCs) of the pristine state (red dashed line)



and the confined state (black dashed line) measured along the $\overline{\Gamma M}$ high symmetry direction and visualized as the second derivative of the photoemission intensity for three different coverages. The second derivative was used to enhance features of the EDC. Both EDCs at 0.52 ML and 0.73 ML were acquired at RT, whereas 60 K was the acquisition temperature for 0.55 ML. The vertical white dotted lines refer to the new symmetry points induced by the molecular network. The red and black dashed lines on the top graph indicate the averaged positions of the maxima of both Lorentzian components fitted for all EDCs.
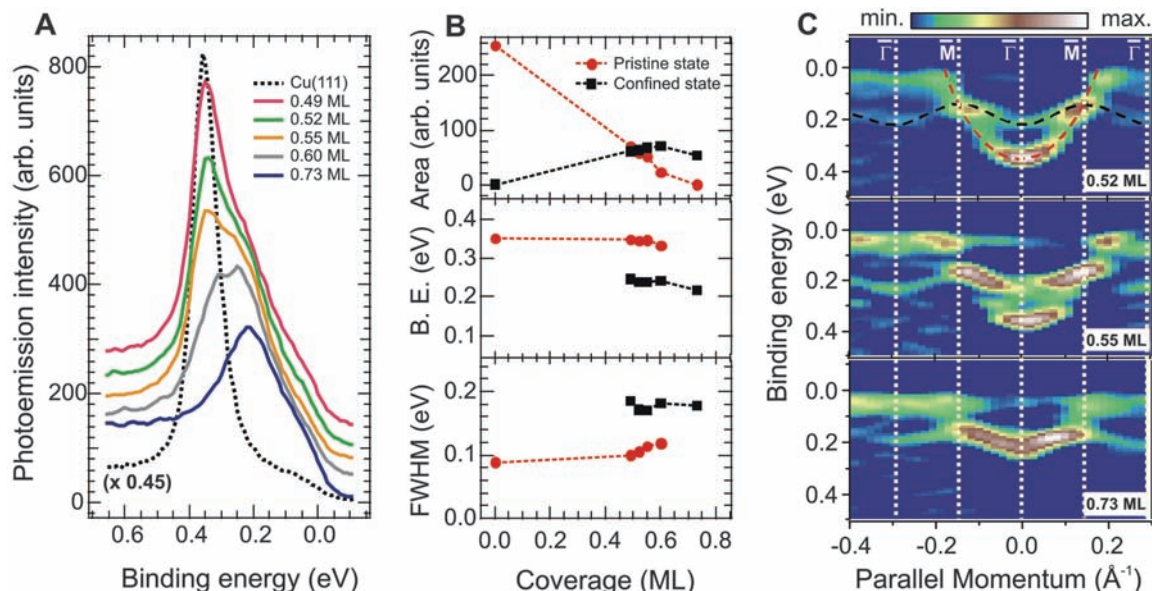
throughout the molecular coverage range, but its intensity decreases as the molecular network coverage increases. It will be referred to below as the pristine state. The component marked black, however, is related to the influence of the molecular network and increases with increasing coverage. In analogy to the pristine case, it has a constant width and a constant binding energy. Its value of $0.23 \pm 0.03$ (experimental error) eV is in very good agreement with the STS spectrum shown in Fig. 2A. Because its binding energy and peak width are independent of the surface coverage, its origin can be attributed to a single pore with its surrounding molecular border, whereas the overall island size shows no influence. This component will be referred to below as the confined state.

Isolated nanocavities exhibit discrete electronic states, but the periodic influence of the porous network on the 2D free electron gas of the surface state, along with the imperfection of its confinement, is expected to induce an electronic band in analogy to the band structure of a solid created by the periodic potential of its atoms. The existence of such cooperative behavior can be investigated by studying with ARPES the electronic structure of both the pristine Cu(111) surface state and the confined state as a function of the surface parallel momentum. The ARPES data displayed in Fig. 3C exhibit for both states dispersive bands. One band (highlighted in red) follows the characteristic parabolic dispersion of the Cu surface state (*25*), whereas the second band (highlighted in black) is related to the periodic potential of the porous network. This interpretation is supported by the periodic continuation of the band within Brillouin zones of higher order that possess the same periodicity as the molecular network (10 times smaller than the substrate). This periodicity

is indicated by white lines in Fig. 3C, which are separated by $|k_{\text{parallel}}| = 0.142$ Å$^{-1}$. Furthermore, weak photoemission intensity very near the Fermi energy and around the $\overline{M}$ symmetry points can also be observed in Fig. 3C, especially for the case of 0.55 monolayer (ML), which was acquired at 60 K. In analogy to the 1D case (*15*, *16*), we attribute this intensity to the existence of the second subband, which has its lowest binding energy at ~60 meV. This value matches perfectly the onset of the second peak observed in the pore STS spectrum of Fig. 2A. Thus, the energy gap between the subbands originating from the confined network is ~90 meV.

The rather shallow dispersing band (bandwidth of ~80 meV) is an indication of the strong confinement of the surface state inside the pores. Further evidence for our interpretation stems from both STS measurements performed on top of dehydro-DPDI molecules and photoemission spectra acquired on a sample fully covered by a close-packed structure of dehydro-DPDI (*21*, *23*). For this non-porous organic network of the same building block, no evidence of the surface state was found, which additionally corroborates the strong confinement of the surface state within the pores. In analogy to the textbook case of electronic states in the solid crystal (*26*), it is the balance of confinement and coupling between neighboring quantum systems that leads to the formation of an electronic band. In our case, the final electronic structure results from the dimensions of the porous network together with the interaction of the molecular backbone with the 2D surface electronic state.

Our results provide conclusive experimental evidence of periodic 2D confinement by a porous molecular network, which can be regarded as a

regular array of 0D quantum dots. A notable consequence of such periodic influence on the otherwise free electron–like surface state is the formation of an artificial electronic band structure. The established and prospective possibilities to control the structures of porous networks, together with the characteristic degree of coupling between ad-molecules and the surface state, will allow the fabrication of related systems with different band structures, resulting in 2D electronic metamaterials in analogy to the well-established optical metamaterials (*27*, *28*). These findings may also provide new insight into the behavior of molecular guests within porous host networks on surfaces (*29*, *30*), and the expected influence of the guests on the electronic band structure may even induce long-range effects in their host-guest behavior (*31*). Moreover, the resulting electronic bands may play a decisive role in the stabilization of the porous networks themselves.

**References and Notes**
1. R. M. Walser, *Proc. SPIE* **4467**, 1 (2001).
2. F. X. Redl, K.-S. Cho, C. B. Murray, S. O'Brien, *Nature* **423**, 968 (2003).
3. A. L. Rogach, *Angew. Chem. Int. Ed.* **43**, 148 (2004).
4. M. F. Crommie, C. P. Lutz, D. M. Eigler, *Nature* **363**, 524 (1993).
5. Y. Hasegawa, P. Avouris, *Phys. Rev. Lett.* **71**, 1071 (1993).
6. L. Gross et al., *Phys. Rev. Lett.* **93**, 056103 (2004).
7. T.-C. Chiang, *Surf. Sci. Rep.* **39**, 181 (2000).
8. M. F. Crommie, C. P. Lutz, D. M. Eigler, *Science* **262**, 218 (1993).
9. J. Kliewer, R. Berndt, S. Crampin, *New J. Phys.* **3**, 22 (2001).
10. J. Li, W.-D. Schneider, R. Berndt, S. Crampin, *Phys. Rev. Lett.* **80**, 3332 (1998).
11. H. Jensen, J. Kröger, R. Berndt, S. Crampin, *Phys. Rev. B* **71**, 155417 (2005).
12. Y. Pennec et al., *Nat. Nanotechnol.* **2**, 99 (2007).
13. U. Bischler, E. Bertel, *Phys. Rev. Lett.* **71**, 2296 (1993).
14. O. Sánchez et al., *Phys. Rev. B* **52**, 7894 (1995).
15. A. Mugarza, J. E. Ortega, *J. Phys. Condens. Matter* **15**, S3281 (2003).

16. J. Lobo *et al.*, *Phys. Rev. Lett.* **93**, 137602 (2004).
17. C. P. Collier, R. J. Saykally, J. J. Shiang, S. E. Henrichs, J. R. Heath, *Science* **277**, 1978 (1997).
18. J. A. Theobald, N. S. Oxtoby, M. A. Phillips, N. R. Champness, P. H. Beton, *Nature* **424**, 1029 (2003).
19. U. Schlickum *et al.*, *Nano Lett.* **7**, 3813 (2007).
20. N. Herron, *J. Inclusion Phenom. Macrocyclic Chem.* **21**, 283 (1995).
21. M. Stöhr *et al.*, *Angew. Chem. Int. Ed.* **44**, 7394 (2005).
22. One monolayer (ML) refers to a full coverage of a close-packed assembly of DPDI. A full surface coverage of the porous network occurs at a coverage of ~0.70 ML[21].
23. Materials and methods are available as supporting material on *Science* online.
24. R. Wiesendanger, Ed., *Scanning Probe Microscopy and Spectroscopy: Methods and Applications* (Cambridge Univ. Press, Cambridge, 1994).
25. F. Reinert, G. Nicolay, S. Schmidt, D. Ehm, S. Hüfner, *Phys. Rev. B* **63**, 115415 (2001).
26. N. Ashcroft, D. Mermin, Eds., *Solid State Physics* (Saunders College, New York, 1976).
27. R. A. Shelby, D. R. Smith, S. Schultz, *Science* **292**, 77 (2001).
28. D. R. Smith, J. B. Pendry, M. C. K. Wiltshire, *Science* **305**, 788 (2004).
29. M. Stöhr, M. Wahl, H. Spillmann, L. H. Gade, T. A. Jung, *Small* **3**, 1336 (2007).
30. H. Dil *et al.*, *Science* **319**, 1824 (2008).
31. A. Kiebele *et al.*, *ChemPhysChem* **7**, 1462 (2006).
32. We thank H.-J. Güntherodt for continuous support; J. Osterwalder and T. Greber for fruitful discussions and providing access to the ARPES experimental station, hosted at the Surface and Interface Spectroscopy beamline of the Swiss Light Source, Paul Scherrer Institut, Villigen, Switzerland; and Nanonis for fruitful collaboration on the STM/STS data acquisition system. Supported by the Swiss National Science Foundation, the National Center of Competence in Research "Nanoscale Science," the European Union through the Marie Curie Research Training Network PRAIRIES, the Wolfermann-Nägeli-Stiftung, the Bundesministerium für Bildung und Forschung (BMBF) Cluster "Forum Organic Electronics" (Heidelberg), and the Swiss Federal Commission for Technology and Innovation (CTI).

# CH Stretching Excitation in the Early Barrier F + CHD₃ Reaction Inhibits CH Bond Cleavage

Weiqing Zhang,[1,2] Hiroshi Kawamata,[1] Kopin Liu[1]*

Most studies of the impact of vibrational excitation on molecular reactivity have focused on reactions with a late barrier (that is, a transition state resembling the products). For an early barrier reaction, conventional wisdom predicts that a reactant's vibration should not couple efficiently to the reaction coordinate and thus should have little impact on the outcome. We report here an in-depth experimental study of the reactivity effects exerted by reactant C-H stretching excitation in a prototypical early-barrier reaction, F + CHD₃. Rather counterintuitively, we find that the vibration hinders the overall reaction rate, inhibits scission of the excited bond itself (favoring the DF + CHD₂ product channel), and influences the coproduct vibrational distribution despite being conserved in the CHD₂ product. The results highlight substantial gaps in our predictive framework for state-selective polyatomic reactivity.

Not all forms of energy are equally effective in driving chemical reactions. Several decades of theory and experiment in reaction dynamics culminated in the Polanyi rules (*1*, *2*), which predict that, in reactions of an atom with a diatomic molecule, reactant vibration and translation have different impacts on the rate. When the barrier is located late along the reaction coordinate (i.e., the transition state resembles products more than reactants), the vibration is considered the more effective driver; the reverse is true for reactions with early barriers (*3*). A similar line of thought underlies efforts directed toward mode-selective chemistry in reactions involving polyatomics. It is now well documented, both in the gas phase (*4–14*) and at surfaces (*15–17*), that excitation of different vibrational motions of a polyatomic reactant can exert a profound influence on chemical reactivity. An intuitively appealing picture emerging from these studies is that exciting a vibrational mode with large displacements along a particular reaction coordinate can preferentially promote the system over the barrier of that pathway, leading to mode-dependent reactivity. The atomic-level mechanism governing this selectivity remains elusive, however.

Previous mode-specific studies have focused on reactions with a late barrier, which seems sensible from the perspective of an extended Polanyi rule framework (*10*). In an early-barrier reaction, reactant vibrations are commonly believed not to couple efficiently to the reaction coordinate. Contrary to this current perception, we report on an experiment that poses serious challenges to our fundamental understanding, even in a qualitative sense, of the vibrational effect on reactivity.

The reaction of F + CHD₃ is highly exothermic, enthalpy of reaction ~ −31 kcal/mol, and, based on Hammond's postulate (*1*), can be regarded as a prototypical early-barrier reaction. Consistent with this characterization, a recent, accurate ab initio calculation of the global potential energy surface indicates a reactant-like transition state structure, Fig. 1, left (*18*). Experimentally, both the thermal rate constant data (*19*) and the crossed-beam scattering results (*20*) suggest a small (<0.4 kcal/mol) barrier to reaction. Previously, comprehensive crossed-beam investigations of the ground state reaction have revealed roughly equal branching to two isotopic product channels, HF + CD₃ and DF + CHD₂ (*20*); numerous rovibrational product states were populated in each channel (*21–23*). We explored the impact on this reaction of one quantum excitation of the C-H stretching vibration ($v_1 = 1$) of CHD₃.

On the basis of the extended Polanyi rule framework described above, we considered several qualitative predictions of the outcome. Given the early barrier, would the vibrational excitation increase the rate relative to the ground state reaction at fixed translational energy? The previously observed equal branching to HF and DF products in the ground state reaction was a clear sign of nonstatistical behavior (*20*). Would initial deposition of vibration energy directly into the C–H bond facilitate its breakage to form HF + CD₃ preferentially? Lastly, the localized nature of the C-H stretch mode suggests that this bond should act as a spectator if the F atom were to attack the D atoms. In keeping with this spectator model (*5*), would the local CH stretching motion retain its vibrational character during the reaction, favoring the CHD₂ product with one-quantum excitation of the CH stretching mode ($v_1 = 1$) over ground state CHD₂ ($v = 0$)?

To seek answers to these questions, we performed a crossed-beam scattering experiment under single-collision conditions. The experimental apparatus consisted of two rotatable pulsed molecular beams and a fixed detector assembly housed in a vacuum chamber (*23*, *24*). The F atom beam was generated by a pulsed high-voltage discharge of 5% F₂ seeded in a pulsed supersonic expansion of Ne at 6 atm. The CHD₃ beam was also produced by pulsed supersonic expansion of 35% CHD₃ seeded in He at 6 atm. Both beams were collimated by double skimmers and crossed in a differentially pumped scattering chamber; collision energy ($E_c$) was tuned by varying the intersection angle of the two molecular beams. A pulsed laser operating near 333 nm probed the nascent distribution of CHD₂

[1]Institute of Atomic and Molecular Sciences (IAMS), Academia Sinica, Post Office Box 23-166, Taipei, 10617 Taiwan. [2]State Key Laboratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Post Office Box 110 extension 11, Dalian 116023, People's Republic of China.

*To whom correspondence should be addressed. E-mail: kliu@po.iams.sinica.edu.tw

and CD$_3$ radicals at the intersection region by (2 + 1) resonance-enhanced multiphoton ionization, REMPI (23), and a time-sliced velocity imaging technique mapped the recoil vector of the CHD$_2^+$ or CD$_3^+$ ions (25). For studies with CH stretch-excited CHD$_3$, an infrared (IR) laser was coupled to a multipass ring-reflector (26) situated just in front of the first skimmer. The narrow band of the IR laser (bandwidth ∼ 0.05 cm$^{-1}$) ensured preparation of CHD$_3$ in a single rovibrational ($v_1 = 1$, $j = 2$) state.

REMPI spectra of the CHD$_2$ products from the F + CHD$_3$ reaction at $E_c = 3.6$ kcal/mol are shown in Fig. 1. The ground state reaction (IR-off) produces predominantly ground state CHD$_2$ (shown as the $0_0^0$ Q head band) and smaller amounts of vibrationally excited CHD$_2$ (the overlapped $2_1^1/3_1^1/5_1^1$ bands) (27, 28). With the IR laser on, the intensities of those bands diminished and a new band $1_1^1$ appeared, signifying the formation of CHD$_2$ ($v_1 = 1$) product. Vibrational excitation induced substantial (45%) depletion of the $0_0^0$ Q head but less depletion (i.e., smaller than $0_0^0$ by about 10%) of the $2_1^1/3_1^1/5_1^1$ bands. To interrogate the influence of CH stretching excitation on other product states, including CHD$_2$ ($4_1^1$), CD$_3$ ($0_0^0$, $1_1^1$, $2_1^1$, $2_2^0$, $3_1^1$, and $4_1^1$), and several combination bands (21–23), we fixed the probe laser wavelengths at the peak of respective bands and recorded the signals alternately for IR-on and IR-off. Surprisingly, they all displayed around 25% IR-associated depletion (when the concurrent $0_0^0$ bands were depleted by 33%, as shown in figs. S1 and S2), suggesting that the yields of all product states except CHD$_2$ ($v_1 = 1$) in the F + CHD$_3$ ($v_1 = 1$) reaction are notably smaller than the corresponding yields in the ground state reaction. This result contrasts sharply with the behavior of the late-barrier Cl + CHD$_3$ ($v_1 = 1$) reaction, where IR irradiation enhanced the formation of vibrationally excited product states that are nearly absent in the ground state reaction (10).

Also overlaid in Fig. 1 are the two IR-on images of the probed CHD$_2$ ($v = 0$) and ($v_1 = 1$) products. Thanks to the time-sliced velocity imaging approach, even the raw data can be easily interpreted by inspection. The energetics of this reaction are well defined, and the vibrational CHD$_2$ product states were selectively detected. By conservation of energy and momentum, the maximum velocities of the DF coproducts recoiling from the selected state of CHD$_2$ were calculated for different final vibrational states and are depicted as dashed lines in Fig. 1. The successive rings in each image can thus be unambiguously assigned to the vibrational states of the dark DF coproduct. The intensity around each ring then gives an immediate impression about the preferred scattering direction of the coincidently formed DF products. Energetically, the initial rovibrational excitation of CHD$_3$ ($v_1 = 1$, $j = 2$) adds 8.63 kcal/mol to the total energy for the IR-on image, and the formation of CHD$_2$ ($v_1 = 1$) products requires at least 8.90 kcal/mol. This near degeneracy (only 0.27 kcal/mol of energy difference) results in energetically similar ring radii of the DF coproduct states in the F + CHD$_3$ ($v = 0$) → CHD$_2$ ($v = 0$) + DF ($v'$) and F + CHD$_3$ ($v_1 = 1$) → CHD$_2$ ($v_1 = 1$) + DF ($v'$) reactions. Remarkably, the angular distributions in the two images are also nearly identical, implicating similar pathways in the two reactions, again at marked variance with the Cl + CHD$_3$ ($v_1 = 1$) reaction, for which the angular distributions differ vastly from that of ground state reaction (9, 10). The $0_0^0$ Q head image with IR-off proved identical to that with IR-on (fig. S1). We therefore concluded that, as predicted, F + CHD$_3$ ($v_1 = 1$) does not produce CHD$_2$ ($v = 0$) and that the change in the $0_0^0$ Q signals upon IR irradiation is due solely to depletion of ground state reactants.

Images acquired for the overlapped $2_1^1/3_1^1/5_1^1$ bands as well as other product states all exhibit very similar features, regardless of whether IR

irradiation is applied (figs. S1 and S2). This result confirms the conjecture from the IR-attenuated REMPI bands that the formation of those product states from the stretch-excited reactant are at most a small fraction of the corresponding product yields in the ground state reaction. On the basis of the degrees of the attenuation of all other CHD$_2$ and CD$_3$ product states, we estimated their collective yields in F + CHD$_3$ ($v_1 = 1$) to be no more than the single product state of CHD$_2$ ($v_1 = 1$). This highly mode-specific behavior is quite unexpected and important. Although enhanced production of CHD$_2$ ($v_1 = 1$) in F + CHD$_3$ ($v_1 = 1$) is anticipated by the spectator model, we did not expect it to be the only detectable product state. The observation of a diminishingly small cross section for the H-atom transfer channel upon C-H stretching excitation, that is, HF < DF, is counterintuitive and exemplifies a strong bond selectivity. We inferred that, upon vibrational excitation, the long-range anisotropic interaction of F atom with a stretched/compressed C-H must change in such a way that it effectively steers the trajectory away from the transition state, practically shutting down the C–H bond scission channel. The mechanism is reminiscent of the stereodynamical effect, induced by the anisotropic van der Waals forces in the reactant valley, that was proposed previously for a preferential formation of DCl over HCl in the Cl + HD ground state reaction (29).

More quantitative information about the predominant CHD$_2$ ($v_1 = 1$) + DF channel can be deduced through image analysis that accounts for density-to-flux transformation (25). Figure 2 summarizes the resultant dynamical attributes from images at $E_c = 1.2$ kcal/mol. Depicted in Fig. 2A are the product speed distributions of the two images, IR-on and IR-off, acquired at the peak of the CHD$_2$ ($1_1^1$) band. From the CHD$_2$ ($0_0^0$) depletion measurement, we found that ∼ 30% of CHD$_3$ reactants were stretch-excited in this case. By scaling down the IR-off distribution by 0.3 and subtracting it from the IR-on data set,

**Fig. 1.** (**Left**) Reaction path energetics for reactant CHD$_3$ initially in the $v = 0$ (black arrow) and $v_1 = 1$ (red) vibrational states. The curve represents schematically a cut through the multidimensional potential energy surfaces governing reactivity. The numbers in the parentheses indicate the vibrational quanta of the product pairs relevant to this study. (**Right**) Two normalized REMPI spectra of the probed CHD$_2$ products, with IR-on (red) and IR-off (black), at $E_c = 3.6$ kcal/mol. Two product images, both with IR-on, are shown for probing of the $1_1^1$Q and $0_0^0$Q bands, respectively. Superimposed on the images are the scattering directions; the 0° angle refers to the initial CHD$_3$ beam direction in the center-of-mass frame.
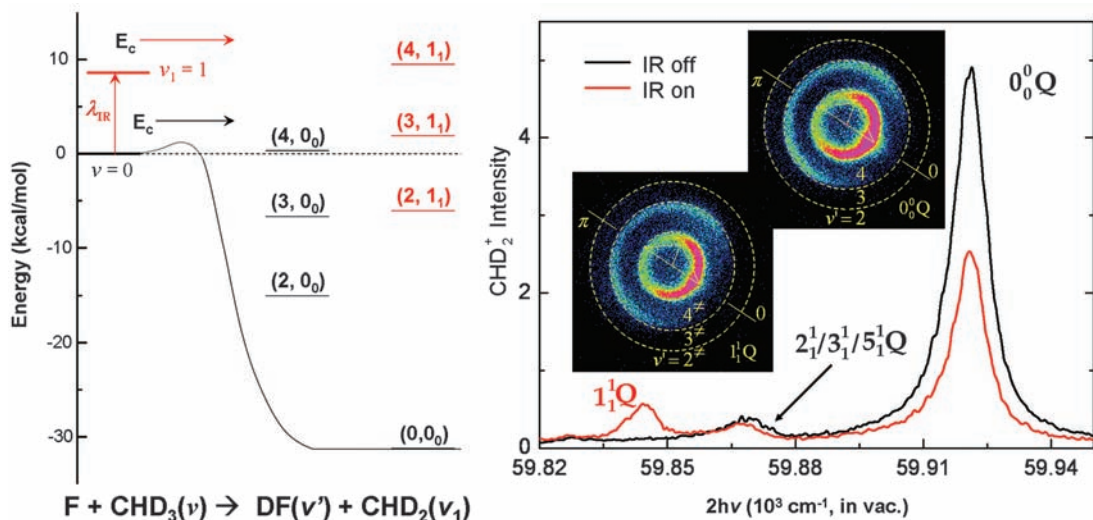
we obtained the genuine distribution for the $CHD_2^{\neq}$ ($v_1 = 1$) product from F + $CHD_3$ ($v_1 = 1$) (the red curve in Fig. 2B). On energetic grounds, the product pair labeled as ($v_{DF}$, $1_1$)$^{\neq}$ can readily be identified. [The superscript $^{\neq}$ indicates the state pair that is produced in the F + $CHD_3$ ($v_1 = 1$) reaction.] Also displayed in Fig. 2B is the speed distribution from the IR-off image acquired at the $CHD_2$ ($0_0^0$) peak (in black); the DF in this

case is vibrationally hotter than in the IR-excited reaction. Such distinct distributions provide compelling evidence that, although the initial C-H stretching excitation of $CHD_3$ transforms adiabatically into the C-H stretching motion of the $CHD_2$ product, it also influences the vibration branching of the DF coproduct. Hence, the locally excited CH bond does not act entirely as a spectator when the F atom attacks the unexcited CD bond.
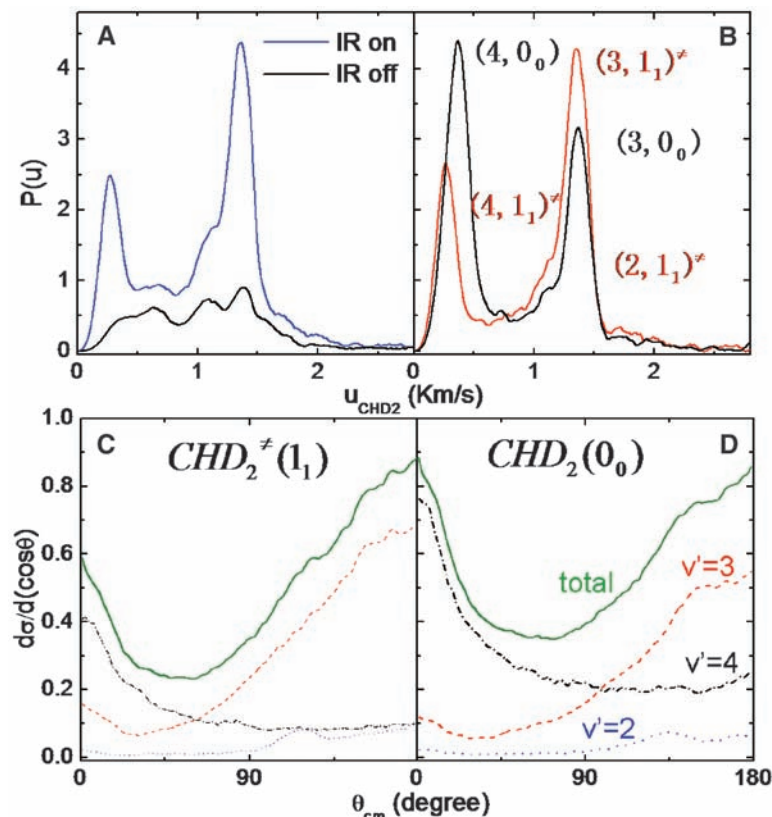


**Fig. 2.** (**A**) The normalized product speed distributions P($\mu$) deduced from the IR-on and IR-off images of the $CHD_2$ ($1_1^1$) band at $E_c = 1.2$ kcal/mol. (**B**) The genuine speed distribution of the F + $CHD_3$ ($v_1 = 1$) → DF ($v$) + $CHD_2$ ($v_1 = 1$) reaction depicted in red after analysis as described in the text. The distribution in black is derived from the IR-off image of the $CHD_2$ ($0_0^0$) band, corresponding to the ground state reaction. On energetic grounds, the peak features can be assigned to the state pairs of the two products, as labeled. The small energetic difference of the two reactions is evident from the slight shift of the peak positions of the (4, $0_0$) and (4, $1_1$)$^{\neq}$ product pairs. (**C** and **D**) Pair-correlated angular distributions from the stretch-excited reaction and ground state reaction, respectively; the quantum number $v'$ refers to the DF coproduct.

**Fig. 3.** Collisional energy dependence of the correlated DF vibrational branching ratio for the ground state reaction (**A**) and the stretch-excited reaction (**B**). The vertical dashed arrows indicate the respective energetic thresholds for forming the DF ($v' = 4$) state.
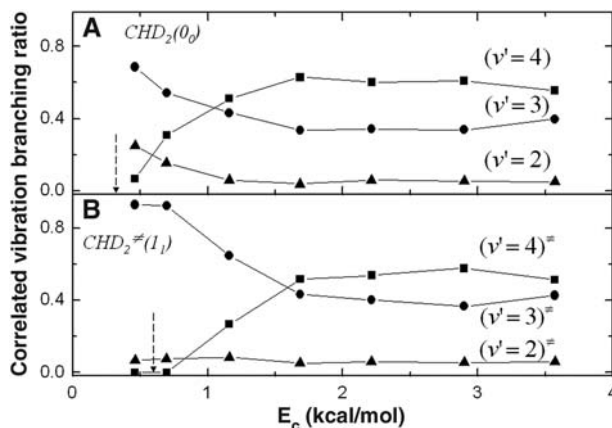


Unlike the speed distributions, the product angular distributions of the two reactions (Fig. 2, C and D) are practically identical (the subtle difference is merely due to the different DF vibrational branching ratios), confirming the initial impression from inspection of the raw images. This contrast in state and angular distributions of the two reactions is instructive. In general the correlated angular distribution, for which the scattering angle should be governed mainly by the trajectories of the two heavy atoms, must also carry the imprint of the initial impact parameter and orientation, that is, a global trait. The correlated DF vibrational distribution, on the other hand, reflects mostly the concerted motions of the transferred D atom and the methyl moiety near the transition state region and is thus more susceptible to local interactions. As such, we reasoned that the differential DF vibrational branching of the two reactions could provide a clue as to how the initial C-H stretching motion affects subsequent events along the reaction trajectory. Because the effects of slight changes in local interaction potential on reactivity tend to more readily manifest at lower collision energies, we examined the $E_c$ dependence of the correlated vibrational branching of DF products in the ground state (Fig. 3A) and stretch-excited (Fig. 3B) reactions. The disparity in the two correlated DF vibrational distributions is substantial for $E_c < 2$ kcal/mol but becomes negligibly small as $E_c$ rises. (Figs. S3 and S4 show the speed and angular distributions, respectively, at higher collision energies.)

Returning to the REMPI spectra shown in Fig. 1, we noted that the enhancement of $CHD_2$ ($1_1^1$) band intensity is smaller than the concurrent depletion of the $CHD_2$ ($0_0^0$) band. Because $CHD_2$ ($v_1 = 1$) is the only product state showing any appreciable formation with IR irradiation, the implication of this discrepancy is that one quantum of CH stretching excitation of $CHD_3$ may actually decelerate the reaction. Repeating the image measurements for both $CHD_2$ ($v_1 = 1$) and $CHD_2$ ($v = 0$) and normalizing them under different collision energies, we obtained the reactive excitation function $\sigma(E_c)$, which is the dependence of the integral cross section on $E_c$, for the ground state and stretch-excited reactants (Fig. 4). To determine the relative cross sections, we exploited an IR-ultraviolet double resonance technique to quantify the relative detection sensitivity $S_{1_1}$ of the $1_1^1$Q and $0_0^0$Q REMPI bands (30) and found that probing the $CHD_2$ ($1_1^1$Q) band is $1.9 \pm 0.7$ times more sensitive than probing the $CHD_2$ ($0_0^0$Q) band. Taking this factor into account, Fig. 4 shows that one quantum excitation of CH stretch mode of $CHD_3$ slows down the reaction rate by about 10-fold and that within our experimental uncertainty the ratio of $\sigma^{\neq}(1_1)/\sigma(0_0)$ in the DF + $CHD_2$ isotope channel exhibits little if any $E_c$ dependency. Because numerous product states are formed abundantly in the ground state reaction, the suppression of overall reactivity by a stretch-excited

**Fig. 4.** (**Top**) Normalized excitation functions of the ground state and stretch-excited reactions. (**Bottom**) Dependence of the ratio of two excitation functions on the collisional energy. The scale on the right ordinate is from the raw results, whereas that on the left gives the true reactivity ratio. The solid curve and the horizontal dashed line are the two possible fits to the experimental results. The vertical bar associated with each data point represents one standard deviation error limit of four to five repeated measurements.
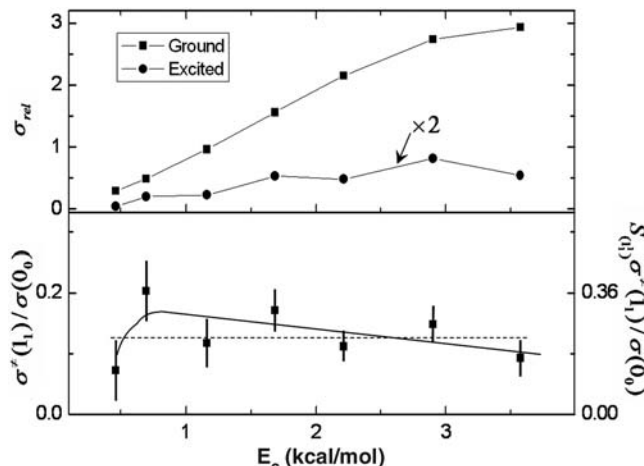


$CHD_3$ reactant is more significant than that shown in Fig. 4.

As alluded to earlier, most studies of vibrational effects on chemical reactivity have focused on late barrier reactions. Several studies of atom plus diatom reactions with early barriers showed vibrational enhancement (*31*–*33*), although extra translational energy accelerated the reactions even more, corroborating the Polanyi rules (*3*, *10*). In contrast, we observed a pronounced negative impact of reactant vibration on total reactivity in F + $CHD_3$ ($v_1 = 1$). Moreover, the initial C-H stretching motion plays a subtle yet decisive role in controlling dynamical attributes. It not only effectively blocks the H atom transfer channel, yielding a counterintuitive product isotope distribution, but also remotely affects the cleavage of the unexcited CD bond, resulting in different pair-correlated distributions in DF + $CHD_2$. In contrast, previous experiments on the Cl + $CHD_3$ ($v_1 = 1$) reaction (*5*, *9*, *10*) demonstrated a rate enhancement by vibration at fixed translational energy, a preferential cleavage of the excited CH bond yielding more HCl than DCl, and the formation of numerous rovibrationally excited $CD_3$ and $CHD_2$ products that are not populated in the ground state reaction. Such a sharp contrast of the reactive behaviors between $CHD_3$ ($v_1 = 1$) with F and $CHD_3$ ($v_1 = 1$) with Cl is particularly illuminating and underscores the pivotal role of the barrier location in directing mode-selective chemistry. Clearly, a conceptual framework of vibrational effects on chemical reactivity is far from complete.

### References and Notes

1. R. D. Levine, R. B. Bernstein, *Molecular Reaction Dynamics and Chemical Reactivity* (Oxford Univ. Press, Oxford, 1987).
2. P. Atkins, J. de Paula, *Atkin's Physical Chemistry* (Oxford Univ. Press, Oxford, ed. 7, 2002), pp. 968–969.
3. J. C. Polanyi, *Science* **236**, 680 (1987).
4. R. N. Zare, *Science* **279**, 1875 (1998).
5. J. P. Camden, H. A. Bechtel, D. J. A. Brown, R. N. Zare, *J. Chem. Phys.* **124**, 034311 (2006).
6. F. F. Crim, *Acc. Chem. Res.* **32**, 877 (1999).
7. F. F. Crim, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12654 (2008).
8. D. L. Proctor, H. F. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12673 (2008).
9. S. Yan, Y.-T. Wu, B. Zhang, X.-F. Yue, K. Liu, *Science* **316**, 1723 (2007).
10. S. Yan, Y.-T. Wu, K. Liu, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12667 (2008).
11. W. T. Duncan, T. N. Truong, *J. Chem. Phys.* **103**, 9642 (1995).
12. J. Palma, J. Echave, D. C. Clary, *Chem. Phys. Lett.* **363**, 529 (2002).
13. Z. Xie, J. M. Bowman, *Chem. Phys. Lett.* **429**, 355 (2006).
14. J. Espinosa-Garcia, J. L. Bravo, C. Rangel, *J. Phys. Chem. A* **111**, 2761 (2007).
15. D. R. Killelea, V. L. Campbell, N. S. Shuman, A. L. Utz, *Science* **319**, 790 (2008).
16. R. D. Beck *et al.*, *Science* **302**, 98 (2003).
17. Z. Liu, L. C. Feldman, N. H. Tolk, Z. Zhang, P. I. Cohen, *Science* **312**, 1024 (2006).
18. G. Czako, B. C. Shepler, B. J. Braams, J. M. Bowman, *J. Chem. Phys.* **130**, 084301 (2009).
19. A. Persky, *Chem. Phys. Lett.* **430**, 251 (2006).
20. J. Zhou, J. J. Lin, K. Liu, *J. Chem. Phys.* **121**, 813 (2004).
21. B. Zhang, J. Zhang, K. Liu, *J. Chem. Phys.* **122**, 104310 (2005).
22. J. Zhou, J. J. Lin, K. Liu, *J. Chem. Phys.* **119**, 8289 (2003).
23. J. Zhou, J. J. Lin, W. Shiu, S.-C. Pu, K. Liu, *J. Chem. Phys.* **119**, 2538 (2003).
24. J. J. Lin, J. Zhou, W. Shiu, K. Liu, *Science* **300**, 966 (2003).
25. J. J. Lin, J. Zhou, W. Shiu, K. Liu, *Rev. Sci. Instrum.* **74**, 2495 (2003).
26. J. Riedel, S. Yan, H. Kawamata, K. Liu, *Rev. Sci. Instrum.* **79**, 033105 (2008).
27. B. Zhang, S. Yan, K. Liu, *J. Phys. Chem. A* **111**, 9263 (2007).
28. In the vibronic band notation used herein, $2_1^1$ designates the REMPI transition involving the $v_2$ vibrational mode with one quantum excitation each in the ground electronic state (the subscript) and the electronically excited state (the superscript).
29. D. Skouteris *et al.*, *Science* **286**, 1713 (1999).
30. W. Zhang, H. Kawamata, A. J. Merer, K. Liu, *J. Phys. Chem. A*, in press; preprint available online at http://pubs.acs.org/doi/full/10.1021/jp902969v.
31. L. J. Kirsch, J. C. Polanyi, *J. Chem. Phys.* **57**, 4498 (1972).
32. J. C. Polanyi, J. L. Schreiber, *Faraday Discuss. Chem. Soc.* **62**, 267 (1977).
33. A. M. G. Ding, L. J. Kirsch, D. S. Perry, J. C. Polanyi, J. L. Schreiber, *Faraday Discuss. Chem. Soc.* **55**, 252 (1973).
34. This work was supported by the National Science Council of Taiwan, Academia Sinica, and the Air Force Office of Scientific Research (grant no. AOARD-09-4030).

# Deep-Sea Temperature and Ice Volume Changes Across the Pliocene-Pleistocene Climate Transitions

Sindia Sosdian*† and Yair Rosenthal

Earth has undergone profound changes since the late Pliocene, which led to the development [~2.7 million years ago (Ma)] and intensification (~0.9 Ma) of large-scale Northern Hemisphere ice sheets, recorded as transitions in the benthic foraminiferal oxygen isotope ($\delta^{18}O_b$) record. Here we present an orbitally resolved record of deep ocean temperature derived from benthic foraminiferal magnesium/calcium ratios from the North Atlantic, which shows that temperature variations are a substantial portion of the global $\delta^{18}O_b$ signal. The record shows two distinct cooling events associated with the late Pliocene (LPT, 2.5 to 3 Ma) and mid-Pleistocene (MPT, 1.2 to 0.85 Ma) climate transitions. Whereas the LPT increase in ice volume is attributed directly to global cooling, the shift to 100,000-year cycles at the MPT is more likely to be a response to an additional change in ice-sheet dynamics.

Two pronounced shifts in the composition of the global benthic foraminifera isotope record ($\delta^{18}O_b$) mark the intensification of Northern Hemisphere glaciation (NHG) during the period from the mid-Pliocene to the late Pleistocene (*1*). The first, starting at ~3.2 to 2.7 million years ago (Ma), marks the late Pliocene transition (LPT) from warm, relatively ice-free conditions to a colder and more glaciated climate in the NH, manifested by increased amplitude of the 41-thousand-year (ky) obliquity cycles in the $\delta^{18}O_b$ record (Fig. 1). The onset of large NH glacial/interglacial cycles is marked by the appearance of ice-rafted debris (IRD) in North Atlantic and North Pacific sediments at 2.7 Ma (*2*, *3*), associated with increased stratification of North Pacific surface

water (4). It has been debated, however, whether these singular events are linked to a threshold response to a shift in background climate state, driven by a decrease in atmospheric carbon dioxide concentration [a decrease in the partial pressure of $CO_2$ ($P_{CO_2}$) (5)], or just a regional expression of a gradual global cooling as suggested by tropical and subtropical records (6, 7). A second shift from ~1.2 to 0.7 Ma, termed the mid-Pleistocene transition (MPT), marks the transition in the dominant periodicity from low-amplitude 41-ky cycles to large-amplitude 100-ky cycles. Mechanisms proposed to explain the MPT include a threshold response to deep ocean cooling and changes in sea-ice growth (8), hypothetically caused by a secular decrease in atmospheric $P_{CO_2}$ (9). An alternative mechanism known as the regolith hypothesis maintains that the exposure of basal rocks under the ice sheets, after long and intense erosion, enabled the buildup of taller ice sheets, independent of changes in background climate (10).

Choosing between the different mechanisms hinges, however, on our ability to separate the $\delta^{18}O_b$ record into its temperature and ice volume components. To date, there is only one orbitally resolved reconstruction of Pliocene-Pleistocene variability in deep ocean temperature (11). The study, based on a ostracode Mg/Ca record from the deep North Atlantic, shows an increase in glacial-interglacial peak-to-peak amplitude [$\Delta$(G-I)] from 3.6°C during the late Pliocene 41-ky glaciations to 4.5°C during the mid- to late Pleistocene 100-ky glaciations, driven mainly by a decrease in glacial temperatures. Because the ostracode record of bottom-water temperature (BWT) covers only the late Pliocene (2.3 to 3.2 Ma) and late Pleistocene [0 to 220 thousand years ago (ka)] intervals, a comprehensive assessment of the evolution of deep ocean temperature, its relationship to NHG, and the transition to the 100-ky G-I cycles has been difficult.

We determined BWT in the deep North Atlantic from Mg/Ca ratios in the benthic foraminifera *Cibicidoides wuellerstorfi* and *Oridorsalis umbonatus*. We used this record to deconvolve changes in the oxygen isotopic composition of seawater ($\delta_\omega$) and estimate ice volume variations as related to changes in deep ocean temperature. The 3.2-million-year (My) record was derived from Deep Sea Drilling Project (DSDP) site 607 (41°N, 32°W; water depth 3427 m) and supplemented with measurements from a nearby piston core (Chain 82-24-23PC; 43°N, 31°W; water depth 3406 m) (Fig. 1) (12). Both sites are currently bathed in North Atlantic Deep Water [NADW, temperature ($T$) = 2.6°C, salinity = 35] and are thus linked to the hydrographic conditions at the

high latitudes (where NADW forms), which are under the direct influence of NH ice sheets. The location and relatively high sedimentation rate at site 607 make it an ideal site for evaluating the role of the deep ocean in Pliocene-Pleistocene climate evolution at orbital-scale resolution (3-ky resolution) (13, 14).

We converted Mg/Ca ratios to paleotemperature using a regional Mg/Ca-temperature calibration (Mg/Ca = 0.15 × BWT + 1.16; fig. S2) that accounts for both temperature and carbonate saturation effects on foraminiferal Mg/Ca (15, 16). It has recently been suggested that the ~0.2 mmol/mol Holocene (HL)–Last Glacial Maximum (LGM) change in Mg/Ca observed in another deep sea core north of site 607 (BOFS 5K) was largely (>70%) driven by a ~20 μmol/kg decrease in $CO_3$ concentration [$CO_3$] (17). In contrast, mid- to late Pleistocene Mg/Ca Δ(G-I) at site 607 varies between 0.4 and 0.6 mmol/mol, and [$CO_3$]

variations may account for only 30 to 40% of the signal. The estimated error of BWT estimation is ±1.1°C (12).

Our temperature record suggests that climate cooling over the past 3.2 My occurred primarily through two distinct events associated with the LPT and MPT shifts in the global $\delta^{18}O_b$ record. BWT variations generally covary and are coherent with the $\delta^{18}O_b$ record in frequency, average trend, and G-I amplitude (Fig. 1). Across the LPT (2.5 to 3.0 Ma), average BWT decreases from 4.5° to 2.5°C, whereas average $\delta^{18}O_b$ increases by 0.71 per mil (‰) (Table 1). In contrast with the ostracode record (11), we showed that the decrease occurs in both interglacial and glacial temperatures. Spectral analysis of the BWT record from 1000 to 2500 ka shows a dominant 41-ky peak, coherent with the $\delta^{18}O_b$ record, signaling the development of obliquity-paced G-I temperature variability after the LPT (fig. S5) (12).
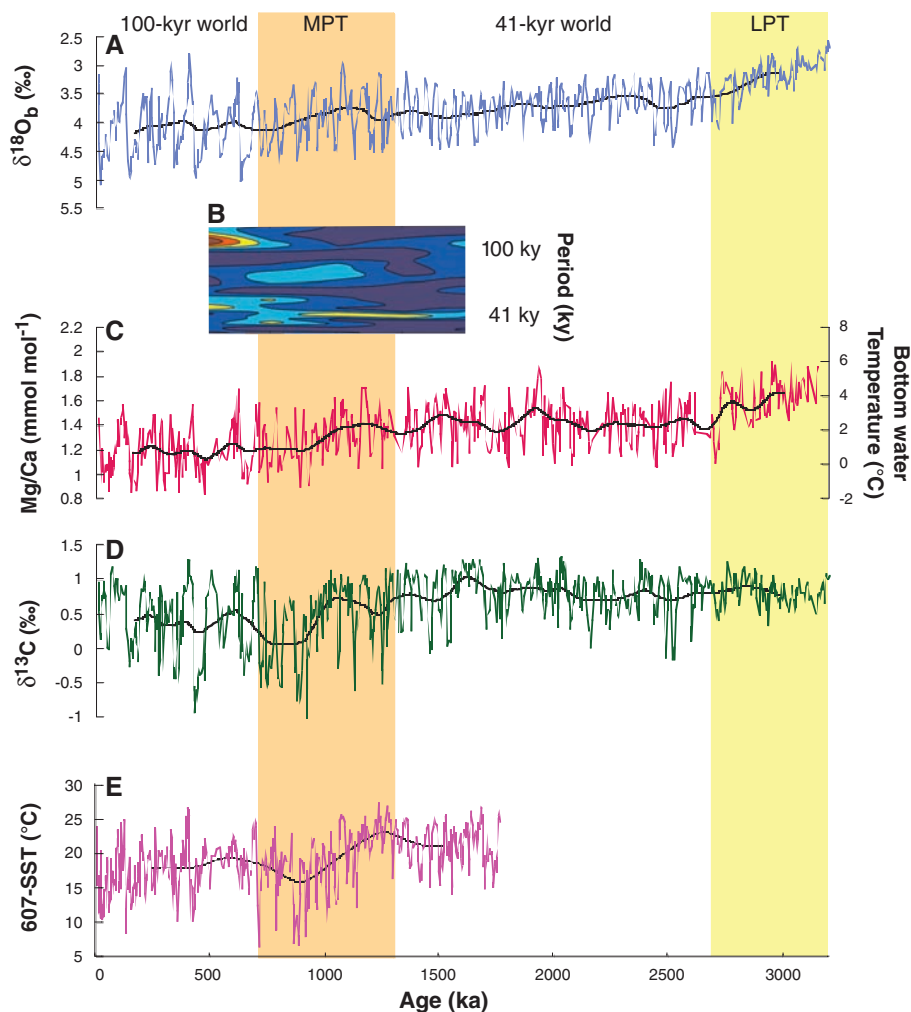


**Fig. 1.** Down-core records from western North Atlantic DSDP site 607. (**A**) Benthic oxygen isotope record [in ‰, using the Pee Dee belemnite (PDB) standard] from (13) and (14). (**B**) Evolutionary spectra of BWT between 500 and 1500 ka, showing that the BWT record is dominated by 41-ky cycles before the MPT and by 100-ky cycles after the transition. (**C**) Benthic foraminiferal Mg/Ca record, derived from *C. wuellerstorfi* and *O. umbonatus*, converted to BWT with the equation Mg/Ca = 0.15 × BWT + 1.16. (**D**) Benthic carbon isotope record ($\delta^{18}O_b$ ‰ PDB) from (13) and (14). (**E**) SST record from DSDP site 607 based on census counts from (13). The MPT and LPT are highlighted by shading. Black lines represent average signals for each variable, determined by applying a Gaussian filter with a cutoff frequency of 400 ky.

Institute of Marine and Coastal Science and Department of Earth and Planetary Sciences, Rutgers University, New Brunswick, NJ 08901, USA.

*To whom correspondence should be addressed. E-mail: sindia.sosdian@anu.edu.au

†Present address: Research School of Earth Sciences, Australian National University, Canberra, ACT, 0200, Australia.

Across the MPT (1150 to 850 ka), average BWT decreases from 2.5° to 1.2°C, whereas average $\delta^{18}O_b$ increases by 0.39‰. As during the LPT, the MPT cooling occurs in both glacial and interglacial temperatures. An evolution spectrum reveals that the 100-ky frequency starts to appear as a broad band around 800 ka but becomes the dominant frequency only at the culmination of the cooling trend, around 700 ka (Fig. 1B). The fact that the cooling and $\Delta$(G-I) amplitude increase precede the frequency shift suggests that either Earth's climate passed a critical threshold or that cooling might not solely have driven the MPT. Subsequently, G-I BWT variability resembles the asymmetric sawtooth pattern of the $\delta^{18}O_b$ record, with gradual cooling into glacial periods followed by abrupt warming during deglaciations.

The BWT record demonstrates that cooling of the deep ocean is a substantial portion of the increase in $\delta^{18}O_b$ at both transitions. For the average long-term temperature trend, cooling accounts for ~70 ± 5% of the positive $\delta^{18}O_b$ shifts across both transitions. During the LPT, BWT $\Delta$(G-I) increases from 2° to 3.3°, whereas $\delta^{18}O_b$ $\Delta$(G-I) increases from 0.6 to 1.2‰, suggesting that on orbital scales, the contribution of temperature to $\delta^{18}O_b$ variability is also ~70 ± 5%. During the MPT, BWT $\Delta$(G-I) increases from 3.3° to 3.9°C, whereas $\delta^{18}O_b$ $\Delta$(G-I) increases from 1.2 to 1.9‰, indicating that temperature accounts for only ~50% of the $\delta^{18}O_b$ $\Delta$(G-I) variability. The decrease in the BWT/ice ratio suggests an increase in the sensitivity of NH ice sheets to changes in global temperature (i.e., more ice per degree of cooling) during the mid- to late Pleistocene.

Cross-spectral analysis between $\delta^{18}O_b$ and BWT reveals that BWT leads $\delta^{18}O_b$ during the late Pliocene to early Pleistocene by 3 ± 2 ky in the obliquity band and by 11 ± 5 ky during the mid- to late Pleistocene in the eccentricity band (Fig. 2). Around 700 ka, the lead increases in association with the appearance of a narrow band of 100-ky variability. The BWT lead in the mid–late Pleistocene is in agreement with other deep ocean temperature studies (11, 15). The phase relationship between BWT and $\delta^{18}O_b$ represents the time constant of ice sheet response to changes in insolation (18). The increase in the phase relationship, coincident with the emergence of 100-ky cycles, indicates that there was a switch in the response time of the NH ice sheets related to a change in their size. Before the MPT, ice sheets were smaller and responded linearly to summer insolation forcing; however, after 700 ka, with the emergence of the 100-ky climate regime, ice sheets grew larger, which delayed their response time to changes in high-latitude temperatures.

The BWT record at site 607 reflects a combination of changes in global temperature and local changes due to water mass variability (13, 14, 19). The $\delta^{13}C_b$ record (Fig. 1D) shows that for the past 3.2 My, site 607 was predominantly bathed during interglacials by relatively warm, salty, and carbonate $[CO_3]$–saturated NADW, mainly reflecting northern Atlantic sea surface conditions. During glacial periods, the site was bathed, to a large extent, by the relatively cold, fresher, and less saturated Antarctic Bottom Water (AABW), which is linked to surface conditions in the Southern Ocean. The relative proportions of NADW ($T$ = 2.6°C) and AABW ($T$ = 0°C), estimated from $\delta^{13}C_b$ (9), vary on both long and orbital time scales. From 3.2 to 2.7 Ma, NADW dominated site 607, reflecting warmer NH polar surface temperature, probably due to strong meridional overturning circulation (20). The decrease in BWT across the LPT occurred in both interglacial and glacial temperatures and thus is attributed mostly to global cooling related to the initiation of NH glaciations. After 2.7 Ma, NADW production decreased but was not completely replaced by AABW, and thus the increase in BWT $\Delta$(G-I) primarily reflects global variability related to the waxing and waning of NH ice sheets. Our record, showing conspicuous cooling during the LPT, contrasts with the ostracode-based reconstruction of (11), which shows no discernible cooling trend across the LPT, and with the results of (7) who, based on an alkenone-derived record of sea surface temperature (SST) from the eastern equatorial Pacific, suggested that climate cooled monotonically from the Pliocene to the late Pleistocene. Here we suggest that our record more accurately reflects a global temperature trend, whereas tropical SST records are also imprinted by regional effects.

From 1150 to 850 ka, the variability in BWT $\Delta$(G-I) reflects high-latitude surface water cooling with a growing influence of water mass variability. After accounting for HL-LGM change in $[CO_3]$, we estimate a $\Delta$(G-I) BWT change of ~3.1°C, which is entirely consistent with the $\Delta\delta^{18}O_b$ change of about 1.5‰ in Chain 82-24-23PC, assuming HL-LGM $\Delta\delta_\omega$ =0.8‰ in the North Atlantic (12). Support for our interpretation of the Mg/Ca record in terms of temperature rather than changes in saturation comes from several lines of evidence. Whereas Mg/Ca shows strong similarity with the $\delta^{18}O_b$ record, it differs significantly in both the general trend and timing of transitions from changes in the deep Atlantic $[CO_3]$ saturation as reflected in the $\delta^{13}C_b$ and $CaCO_3$ preservation records (12). On an orbital scale, BWT leads changes in $\delta^{13}C_b$ (table S2), suggesting that circulation-driven changes in carbonate saturation are not the dominant control of Mg/Ca. We estimate that changes in ocean circulation contributed <0.5°C to the total glacial cooling of 1.3°C, suggesting an average decrease of BWT of about 0.9°C during both glacial and interglacial maxima due to high-latitude cooling. Indeed, the trend, $\Delta$(G-I), and frequency changes are entirely consistent with SST records from site 607 (Fig. 1E), Ocean Drilling Program (ODP) site 982 (21), and tropical upwelling regions, suggesting that we are capturing a global climate shift (13, 22).

We use the paleotemperature equation [$T$ = 16.9 – 4.0 ($\delta^{18}O_b - \delta_\omega$)] to calculate the $\delta^{18}O$ composition of seawater (23). We assume that $\delta_\omega$ variability mainly reflects variations in global ice

**Table 1.** Statistical summary of the primary features of the Mg/Ca-BWT and $\delta_\omega$ record, specifically the mean (M), glacial (G), and interglacial (I) trends and glacial-interglacial $\Delta$(G-I) amplitude from different time intervals.

| Time interval (ky) | Mg/Ca-BWT (°C) | | | | $\delta_\omega$ (‰) | | | |
|---|---|---|---|---|---|---|---|---|
| | M | G | I | $\Delta$(G-I) | M | G | I | $\Delta$(G-I) |
| 0 to 850 | 1.2 | −0.5 | 3.3 | 3.9 | 0.35 | 0.91 | −0.08 | 0.99 |
| 1150 to 2500 | 2.5 | 0.7 | 4.2 | 3.3 | 0.28 | 0.62 | 0.02 | 0.60 |
| 3000 to 3150 | 4.5 | 3.4 | 6.0 | 2.0 | 0.07 | 0.18 | −0.13 | 0.31 |
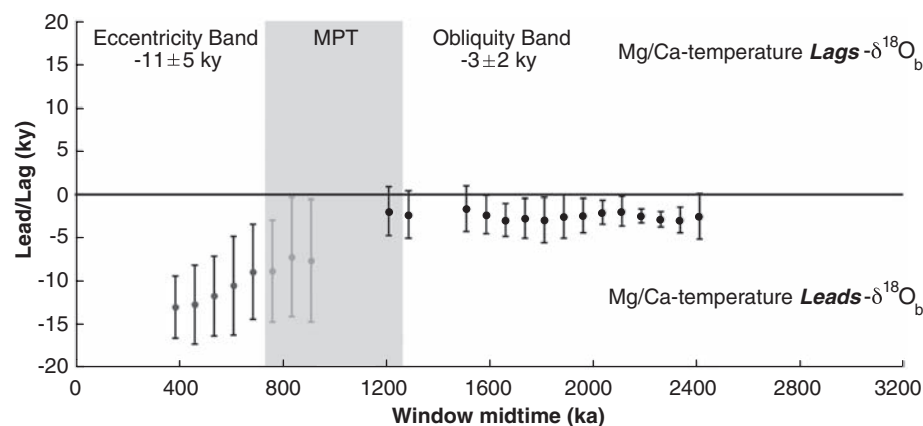


**Fig. 2.** Phase and coherency relation between $\delta^{18}O_b$ and Mg/Ca-BWT at the eccentricity (100-ky) and obliquity (41-ky) bands. Intervals that are coherent at 80% and 95% confidence levels are shown with gray bars and black bars, respectively. Before coherency and phase analysis, all records were interpolated to even intervals of 3-ky resolution. Phases were computed with the use of the ARAND program iterative spectra feature, with a 300-ky window and 250-ky lags.

volume, where a 10-m change in sea level equates to a 0.1‰ change in $\delta_\omega$ (24), which is on the upper side of, though consistent within, uncertainties with estimates of changes in $\delta_\omega$ (25), and yields the best fit with sea level reconstructions over the past 450 ky (Fig. 3) (21). Given an error of ±1.1°C in BWT estimation and an uncertainty of ~0.2‰ in $\delta^{18}O_b$ (12), the uncertainty of sea level reconstruction is ±32 m for $\Delta$(G-I) amplitude and ±21 m for the mean trend (12). The $\delta_\omega$ record shows a gradual increase in continental ice from the mid-Pliocene to the late Pleistocene, primarily due to an increase in ice volume during glacial intervals. Between ~2.7 and 1.2 Ma, average $\Delta$(G-I) sea level fluctuations are about 60 to 80 m, increasing to 120 ± 32 m in the late Pleistocene. It is noteworthy that interglacial ice extent inferred from the $\delta_\omega$ record has been relatively constant since 3.2 Ma, in contrast with the benthic $\delta^{18}O$ record without temperature correction.

Our sea level reconstruction is generally consistent with previous studies suggesting 40 to 70 m of peak-to-peak $\Delta$(G-I) in the late Pliocene and the increase in glacial ice toward the late Pleistocene (11, 12, 26). However, our new record provides critical details necessary for assessing the evolution of NH ice sheets in relation to changes in the climate background state. Our record shows that the LPT cooling is associated with an average sea level drop of ~22 ± 21 m due to 44- and 15-m drops during glacial and interglacial intervals, respectively (Table 1). Within errors, these estimates are consistent with previous estimates of

30 to 35 m of sea level fall (11, 26, 27, ). Across the MPT, glacial sea level drops by ~20 ± 21 m, whereas interglacial sea level stays relatively stable, resulting in an average increase in ice volume of about 8 ± 21 m, associated with a global cooling of ~0.9°C. The glacial estimate is consistent with a reconstruction of low sea level stands, based on depositional sequences from shallow marine sediments, which found a 20 to 30 m decrease from 1000 to 900 ka (28). An additional and comparable drop in glacial sea level occurs in the late Pleistocene, after the transition to 100-ky periodicity. However, our records show that this last interval is also characterized by more extreme deglaciations manifested by higher (15 ± 32 m) interglacial sea level stands. Apparently, after the MPT, less NH ice persisted in interglacial periods despite the cooler climate.

The divergence in sea level and similarity in deep ocean temperature history across the LPT and MPT suggest that each transition may be responding to a different set of forcings. The global cooling of the deep ocean and concomitant increase in ice volume across the LPT may reflect a high-latitude response to a secular decrease in atmospheric $PCO_2$ (5). Our 2°C cooling across the LPT is consistent with modeling studies suggesting that the inception of NHG may be linked to a substantial drop in air temperature during the late Pliocene (29, 30). Deep ocean cooling associated with the MPT has been highlighted as another threshold response to radiative forcing driven by a decrease in atmospheric $PCO_2$, which resulted in

a larger areal extent of NH ice sheets. This would have led to larger fluctuations of planetary albedo, resulting in enhanced BWT $\Delta$(G-I) and sea level amplitudes due to colder and "icier" glacials and warmer, less icy interglacials. Alternatively, the regolith hypothesis maintains that changes in basal rock conditions enabled the buildup of taller (rather than expanded) ice sheets, independent of G-I changes in albedo (10). Our record shows that deep ocean cooling from 1150 to 825 ka precedes the major expansion of ice sheets and the frequency shift from 41-ky to 100-ky glaciations at 700 ka, thereby suggesting that although cooling might have played an important role, it was not sufficient to explain the MPT. Two lines of evidence support this hypothesis. First, the decrease in G-I BWT–to–ice volume ratio during the MPT suggests an increase in the sensitivity of NH ice sheets to changes in global temperature, which cannot be directly attributed to increased albedo. Second, although the glacial increase in ice volume suggested by the $\delta_\omega$ record is consistent with the cooling of glacial intervals, the interglacial decrease of ice volume is inconsistent with the cooling of interglacial intervals. Both observations point to a fundamental change in ice-sheet dynamics rather than just a threshold response to global cooling. This corollary is consistent with the fact that the frequency change occurs after the major cooling interval.

Our data support the hypothesis that the MPT represents a fundamental change in ice-sheet dynamics that is consistent with the growth of thicker, more unstable ice sheets that fully deglaciate during interglacial periods (10, 30). The increased abundance of IRD after the MPT is consistent with an increase in ice-sheet thickness (31) and with geological evidence suggesting that ice sheets grew in height rather than areal extent (10). We conclude that the cooling associated with the MPT might have preconditioned the NH to allow the growth of taller glacial ice sheets but cannot account for the shift in frequency or sea level history.
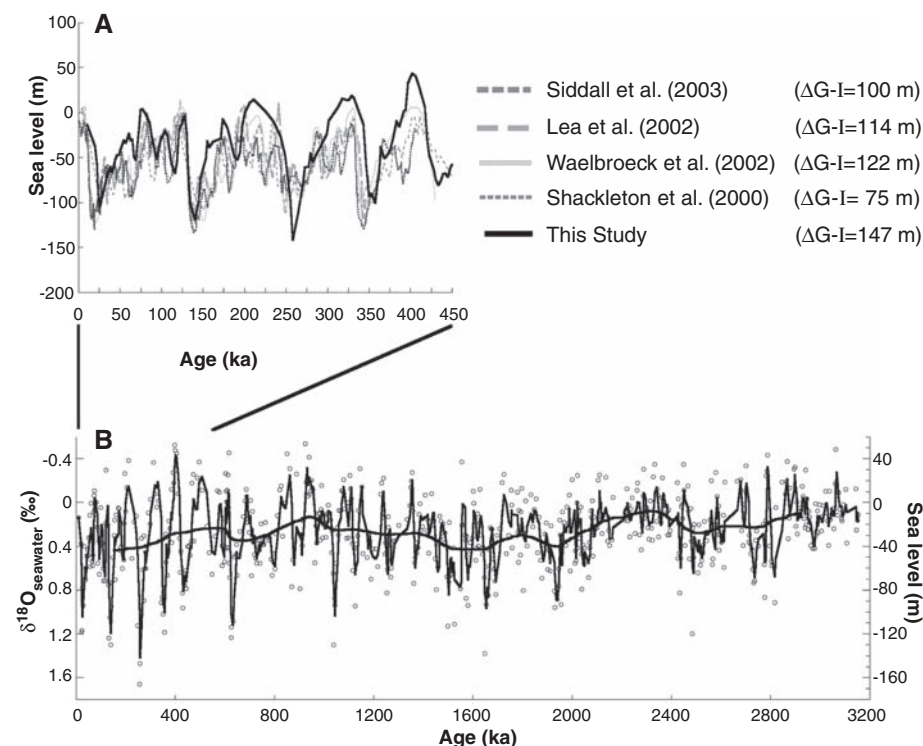


**Fig. 3.** (**A**) Sea level record from 0 to 450 ka derived from Mg/Ca-BWT estimates and $\delta^{18}O_b$ from Chain 82-24-23PC and DSDP site 607, compared with other published sea level records (17, 32–34). (**B**) $\delta_\omega$ record from 0 to 3.2 Ma (‰, using standard mean ocean water). The solid block line in (A) and (B) represents a three-point smoothed curve of the $\delta_\omega$ record (12).

**References and Notes**
1. N. J. Shackleton, N. D. Opdyke, *Quat. Res.* **3**, 39 (1973).
2. N. J. Shackleton *et al.*, *Nature* **307**, 620 (1984).
3. M. A. Maslin, X. S. Li, M.-F. Loutre, A. Berger, *Quat. Sci. Rev.* **17**, 411 (1998).
4. D. M. Sigman, S. L. Jaccard, G. H. Haug, *Nature* **428**, 59 (2004).
5. W. F. Ruddiman, M. E. Raymo, in *The Past Three Million Years: Evolution of Climatic Variability in the North Atlantic Region* (Cambridge Univ. Press, Cambridge, 1988), pp. 227–234.
6. A. C. Ravelo, D. H. Andreasen, M. Lyle, A. O. Lyle, M. W. Wara, *Nature* **429**, 263 (2004).
7. K. T. Lawrence, Z. Liu, T. D. Herbert, *Science* **312**, 79 (2006).
8. E. Tziperman, H. Gildor, *Paleoceanography* **18**, 10.1029/2001PA000627 (2003).
9. M. E. Raymo, D. W. Oppo, W. B. Curry, *Paleoceanography* **12**, 546 (1997).
10. P. U. Clark, D. Pollard, *Paleoceanography* **13**, 1 (1998).
11. G. S. Dwyer *et al.*, *Science* **270**, 1347 (1995).
12. Materials and methods are available as supporting material on *Science* Online.
13. W. F. Ruddiman, M. E. Raymo, D. G. Martinson, B. M. Clement, J. Backman, *Paleoceanography* **4**, 353 (1989).

14. M. E. Raymo, W. F. Ruddiman, J. Backman, B. M. Clement, D. G. Martinson, *Paleoceanography* **4**, 413 (1989).
15. P. A. Martin *et al.*, *Earth Planet. Sci. Lett.* **198**, 193 (2002).
16. H. Elderfield, J. Yu, P. Anand, T. Kiefer, B. Nyland, *Earth Planet. Sci. Lett.* **250**, 633 (2006).
17. J. Yu, H. Elderfield, *Earth Planet. Sci. Lett.* **276**, 129 (2008).
18. N. J. Shackleton, *Science* **289**, 1897 (2000).
19. E. A. Boyle, L. D. Keigwin, *Earth Planet. Sci. Lett.* **76**, 135 (1985).
20. M. E. Raymo, B. Grant, M. Horowitz, G. H. Rau, *Mar. Micropaleontology* **27**, 313 (1996).
21. K. T. Lawrence, T. D. Herbert, C. M. Brown, M. E. Raymo, A. M. Haywood, Paleoceanography 24, PA2218, 10.1029/2008PA001669 (2009).
22. E. L. McClymont, A. Roselle-Mele, *Geology* **33**, 389 (2005).
23. N. J. Shackleton, *Cent. Natl. Rech. Sci. Colloq. Int.* **219**, 203 (1974).
24. R. G. Fairbanks, *Nature* **342**, 637 (1989).
25. J.-C. Duplessy, L. Labeyrie, C. Waelbroeck, *Quat. Sci. Rev.* **21**, 315 (2002).
26. M. Mudelsee, M. E. Raymo, *Paleoceanography* **20**, 10.1029/2005PA001153 (2005).
27. T. M. Cronin *et al.*, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **108**, 437 (1994).
28. A. Kitamura, T. Kawagoe, *Quat. Sci. Rev.* **25**, 323 (2006).
29. D. J. Lunt, G. L. Foster, A. M. Haywood, E. J. Stone, *Nature* **454**, 1102 (2008).
30. R. Bintanja, R. S. W. van de Wal, *Nature* **454**, 869 (2008).
31. D. A. Hodell, J. E. T. Channell, J. H. Curtis, O. E. Romero, U. Rohl, *Paleoceanography* **23**, PA4218, 10.1029/2008PA001591 (2008).
32. C. Waelbroeck *et al.*, *Quat. Sci. Rev.* **21**, 295 (2002).
33. M. Siddall *et al.*, *Nature* **423**, 853 (2003).
34. D. W. Lea, P. A. Martin, D. K. Pak, H. J. Spero, *Quat. Sci. Rev.* **21**, 283 (2002).

# Transient Simulation of Last Deglaciation with a New Mechanism for Bølling-Allerød Warming

Z. Liu,[1,2,3]* B. L. Otto-Bliesner,[4] F. He,[3] E. C. Brady,[4] R. Tomas,[4] P. U. Clark,[5] A. E. Carlson,[6] J. Lynch-Stieglitz,[7] W. Curry,[8] E. Brook,[5] D. Erickson,[9] R. Jacob,[10] J. Kutzbach,[3] J. Cheng[1,3]

We conducted the first synchronously coupled atmosphere-ocean general circulation model simulation from the Last Glacial Maximum to the Bølling-Allerød (BA) warming. Our model reproduces several major features of the deglacial climate evolution, suggesting a good agreement in climate sensitivity between the model and observations. In particular, our model simulates the abrupt BA warming as a transient response of the Atlantic meridional overturning circulation (AMOC) to a sudden termination of freshwater discharge to the North Atlantic before the BA. In contrast to previous mechanisms that invoke AMOC multiple equilibrium and Southern Hemisphere climate forcing, we propose that the BA transition is caused by the superposition of climatic responses to the transient $CO_2$ forcing, the AMOC recovery from Heinrich Event 1, and an AMOC overshoot.

The last deglaciation (~21 to 11 ka) (ka: 1000 years ago) experienced the last major natural global warming and was punctuated by several abrupt climate changes (*1*, *2*). Particularly notable changes occurred in the North Atlantic region where the surface climate experienced cooling during Heinrich Event 1 (H1, ~17 ka), followed by an abrupt warming at the onset of the Bølling-Allerød (BA, ~14.5 ka) (Fig. 1) (*2–11*). These abrupt climate changes were accompanied by large changes in the Atlantic meridional overturning circulation (AMOC), suggesting a causal linkage through the AMOC and its associated heat transport (*5*, *6*) (Fig. 1C).

Climate evolution during the last deglaciation has been studied in transient simulations with climate models of intermediate complexity (*12*). In particular, some intermediate models simulated abrupt warming events like the BA by triggering a resumption of the AMOC either locally by a reduced meltwater flux (MWF) (*13*) or surface warming (*14*) over the North Atlantic, or remotely by an increased MWF (*15*) or surface warming (*16*) over the Southern Ocean. In all the cases, the abrupt warming occurred in response to a gradually varying forcing (*17*) through a strong hysteresis associated with AMOC advection and North Atlantic convection (*18*). Long transient simulations, however, have not been carried out in synchronously coupled atmosphere-ocean general circulation models (CGCMs), which

include the most advanced climate physics and are currently being used for future climate projections. Here we present a transient simulation of the climate evolution from the Last Glacial Maximum (LGM, ~21 ka) to BA using a state-of-art CGCM: the National Center for Atmospheric Research Community Climate System Model version 3 (NCAR CCSM3) (*19*). Through realistic changes in boundary conditions and forcing, our simulation captures many major features of the deglacial climate evolution, including the magnitude of the climate response as inferred from observations.

Starting from a previous LGM simulation (*20*), our model was integrated from 22 toward 14 ka, forced by changes in insolation (*21*), atmospheric greenhouse gas (GHG) concentrations (*22*) (Fig. 1A), continental ice sheets and coastlines (*23*, *24*), and MWF over the North Atlantic and Gulf of Mexico (Materials and Methods 1). From 22 to 19 ka, the model climate changes slowly, primarily due to insolation forcing. The simulated Atlantic Ocean at 19 ka, which we call the glacial state, captures important features of the LGM circulation as reconstructed from various proxy records (*25*, *26*), including a shallower North Atlantic Deep Water (NADW) accompanied by a southward shift of deep convection from the Nordic Seas to the Greenland Sea, a reduced AMOC transport, and a volumetric expansion of Antarctic Bottom Water (Fig. 2, A and D and figs. S1A and S2) (Materials and Methods 1).

From 19 to 17 ka, we applied a MWF derived from Northern Hemisphere ice sheets to the North Atlantic and Gulf of Mexico at a rate consistent (within uncertainties) with the record of sea-level rise (*23*, *27*), gradually reaching a peak flux of 20 m per thousand years (ky) at H1 (Fig. 1, B and C). The MWF was then reduced in two scenarios: a linear decrease to zero at 14.2 ka (DGL-B) and a constant flux (of 15 m/ky) until a sudden shut-off at 14.67 ka (DGL-A) (Materials and Methods 2). Because the meltwater termination scenarios DGL-B and DGL-A represent the slowest and fastest possible MWFs, the two corresponding experiments represent two

[1]Key Laboratory of Meteorological Disaster, Nanjing University of Information Science and Technology, Nanjing, 210044, China. [2]State Key Laboratory of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an, 710075, China. [3]Center for Climatic Research and Department of Atmospheric and Oceanic Sciences, University of Wisconsin, Madison, WI 53706, USA. [4]Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, CO 80307–3000, USA. [5]Department of Geosciences, Oregon State University, Corvallis, OR 97331, USA. [6]Department of Geology and Geophysics and Center for Climatic Research, University of Wisconsin, Madison, WI 53706, USA. [7]School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA. [8]Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA. [9]Oak Ridge National Laboratory, Oak Ridge, TN 37831–6016, USA. [10]Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439–4843, USA.

*To whom correspondence should be addressed. E-mail: zliu3@wisc.edu

end members for simulations under more realistic MWF.

The increase in MWF starting at 19 ka induces a gradual decrease in the AMOC (Fig. 1D). The associated freshwater anomaly is confined initially to the upper North Atlantic at H1 (fig. S1B) and is then transported in the upper ocean into the Southern Ocean, where it eventually spreads northward in the deep ocean, substantially freshening the glacial bottom water by the time of the BA (fig. S1C). From 17 ka to the BA, the MWF decreases in both experiments, leading to increases in the AMOC. The AMOC increases gradually toward the BA following the gradual decrease in MWF in DGL-B, but it

resumes abruptly at the BA following the abrupt termination of MWF in DGL-A (Fig. 1D). Regardless of the recovery speeds, however, the AMOC in both experiments peaks at ~19 sverdrup (1 sverdrup = $10^6$ m$^3$/s) at the onset of the BA, or ~6 sverdrup greater than the glacial-state transport (~13 sverdrup), and is characterized by a deeper and stronger circulation (Fig. 2C), comparable with that in a Holocene simulation (not shown). The simulated transient responses of the AMOC from 19 to 14 ka, especially in experiment DGL-A, are in overall agreement with a reconstruction of changes in the AMOC export (Fig. 1D) (5).

Accompanying these changes in the AMOC is a bipolar seesaw response in surface temper-

ature until H1, followed by a global warming that peaks at the BA. The bipolar seesaw response is characterized by a cooling over the Northern Hemisphere and a warming over the Southern Hemisphere (Fig. 3A) and is caused by a decrease in the northward heat transport of the AMOC (28–30). The suppression of convection also contributes to the strong surface cooling in the North Atlantic. A weak warming also occurs over North America and northern Europe due to the lowering of the ice sheets and the associated response of the planetary waves (31). By contrast, the warming from H1 to BA is global, with the maximum warming relative to H1 exceeding 20°C in the North Atlantic and Arctic (Fig. 3B). A large fraction of this BA warming, or that part relative to the glacial state (Fig. 3C), is characterized by a polar amplification in both hemispheres.

The simulated global temperature evolution closely resembles paleoclimate reconstructions [supporting online material (SOM) Text 1], although there is a tendency for a model-data discrepancy before 19 ka, which may be attributed partly to our initial state of an equilibrium climate at LGM. Focusing on the Atlantic sector here, the simulated annual temperature closely follows the trajectory of temperature reconstructions from Greenland (Fig. 1E) and Antarctic (Fig. 1F) ice cores. From 21 to 19 ka, both Greenland and Antarctica show a weak early warming of ~0.5°C, which we attribute to increased obliquity and associated sea-ice feedback (32, 33). With the increased MWF from 19 to 17 ka, temperature decreases by 4°C over Greenland but increases by 2°C over Antarctica, reflecting the bipolar seesaw response (Fig. 3A). The most pronounced changes occur in response to the decrease in MWF from 17 to 14 ka. Simulated Greenland temperature increases by 15°C at the BA onset, comparable with temperature reconstructions (7, 34). By contrast, Antarctic temperature continues to increase toward the BA, as in ice-core reconstructions (35), which is caused primarily by the rapid increase in GHG concentrations during this period (SOM Text 2).

The characteristic North Atlantic temporal evolutionary structure of H1 cooling followed by BA warming simulated by CCSM3 is also in good agreement with sea-surface temperature (SST) reconstructions from the eastern subtropical gyre off the Iberian Margin (Fig. 1G). Over the tropical Atlantic, model SSTs first decrease toward H1 and then recover sharply at the BA, accompanied by a suppression of rainfall toward the H1 and a subsequent recovery toward the BA. The simulated rainfall suppression toward H1 (enhancement at BA) is caused by the southward (northward) migration of the Intertropical Convergence Zone (ITCZ), which is induced by the surface ocean warming south (north) of the equator in response to the increased (decreased) freshwater forcing (36). The simulated abrupt increase of SSTs and rainfall at the onset of the BA, especially in experiment DGL-A, generally



**Fig. 1.** Data-model comparison for several benchmark time series. (**A**) June insolation at 60°N (purple) (21) and atmospheric $CO_2$ concentration (green) (22). ppmv, parts per million by volume. (**B**) Sea level from the reconstruction (gray) (23) and model [meters of equivalent global sea level (ESL) for meltwater]. (**C**) Freshwater fluxes (FWF) in the model. (**D**) Pa/Th ratio at Bermuda (GGC5 core) as a proxy for AMOC export (5), and model maximum AMOC transport (below 500 m). (**E**) Greenland surface air temperature (SAT) based on Greenland Ice Sheet Project 2 (GISP2) $\delta^{18}O$ reconstruction with borehole temperature calibration (34) and in the model (model offset by −3°C). (**F**) Antarctic surface air temperature based on Dome C $\delta^{18}O$ reconstruction (35) and in the model. (**G**) SST from the Iberian Margin from reconstructions (8, 9) and model. (**H**) SST from the Cariaco Basin from reconstruction (10) and model (model offset by 4°C). (**I**) Rainfall in Cariaco Basin from reconstruction (11) and model. In (B) to (I), gray is used for the reconstruction, and red and blue for experiments DGL-A and DGL-B, respectively. The five circles on DGL-A in (D) represent the glacial state (GLA, 19 ka), H1 (17 ka), PreBA (14.7 ka), Recovery (REC, 14.5 ka), and BA (14.35 ka). All model variables are annual means with a 20-year running average. Overall, model simulations, especially DGL-A, are in good agreement with the proxy records, especially outside the tropical Atlantic. BP, before present.

agree with the Cariaco Basin reconstructions (Fig. 1, H and I), although the simulated earlier decrease in the SST and rainfall toward H1 seems to be largely absent in the reconstructions. Overall, the model overestimates the climate variability associated with the ITCZ over the tropical Atlantic. This overestimation is likely caused by a double ITCZ bias in the model tropical Atlantic climatology—a common deficiency in most current CGCMs (37).

In contrast to the bipolar seesaw temperature response at the ocean surface, the subsurface ocean warms throughout the Atlantic during the MWF period from H1 to the BA (Fig. 2, E and F), providing a heat reservoir potentially important for the subsequent BA warming. The subsurface warming is largely consistent with previous experiments and observations (38). In the North Atlantic, the surface freshening suppresses the convective heat exchange, cooling the surface but warming the subsurface. The subsequent reduction of the AMOC and associated northward heat transport warms the entire South Atlantic water column, but further cools the surface North Atlantic.

Previous studies of the BA in simplified models found that the abrupt warming was caused by a sudden resumption of the AMOC in response to a gradual perturbation forcing, reflecting a strongly nonlinear response to MWF associated with substantial AMOC hysteresis (13–16). By contrast, CCSM3 simulates the BA warming largely as a linear response to MWF, with most of the abrupt warming occurring only in response to a sudden termination of the MWF. Indeed, CCSM3 has no appreciable hysteresis (SOM Text 3). This is best illustrated in experiment DGL-B from 19 to 14 ka (39, 40), whereby the AMOC decreases gradually from 19 to 17 ka when the MWF increases gradually, but then recovers gradually from 17 to 14.2 ka when the MWF decreases slowly, eventually overshooting beyond the glacial transport (Fig. 1, C and D) (41).

Both experiments DGL-A and DGL-B simulate a ~15°C warming over Greenland from H1 to BA, comparable with temperature reconstructions (7, 34) (Fig. 1). Of this amount, 5°C is associated with the AMOC recovery from H1 back to the glacial state, and the remaining 10°C results from the $CO_2$-induced warming and an AMOC overshoot (the AMOC recovery beyond the glacial-state transport). The approximate contributions of the three mechanisms can be assessed from experiment DGL-A. First, the total radiative warming due to $CO_2$ and orbital forcing in the absence of AMOC change can be estimated as the difference between the pre-BA (14.67 ka) and H1 states. In this period, the MWF and, in turn, the collapsed AMOC remain unchanged in DGL-A (Fig. 1, C and D), so that the warming is caused primarily by the radiative forcing. Indeed, from H1 to pre-BA, both Greenland and Antarctica exhibit a similar gradual warming of ~4°C (Fig. 1, E and F). This symmetric warming (fig. S6A) is consistent with the symmetric radiative forcing of $CO_2$ and annual

insolation. Our further sensitivity experiments show, however, that this warming is dominated by the $CO_2$ forcing, whereas the contribution of the orbital forcing is weak (SOM Text 2). Indeed, the period from 17 to 14.6 ka saw the first major rise of the GHGs, with a 40-ppm (parts per million) increase in $CO_2$ accounting for about half of the glacial-interglacial change (Fig. 1A). This large increase in $CO_2$ induces a symmetric global warming background for the subsequent BA warming.

Upon the suspension of the MWF after 14.67 ka in experiment DGL-A, the AMOC recovers rapidly to its glacial level by ~14.5 ka, with a 5°C warming over Greenland (Fig. 1, D and E). This recovery warming is of the same magnitude as the H1 cooling, reflecting a nearly linear dependence of Greenland temperature on AMOC strength. In contrast to the bipolar response to the MWF during H1 (Fig. 3A), however, the rapid warming due to AMOC recovery is confined to the North Atlantic and Arctic regions (Fig. 1, E to H, and fig. S6B).

Finally, after the recovery at 14.5 ka, Greenland temperature increases by another 6°C, peaking at 14.35 ka (Fig. 1E), accompanied by an AMOC overshoot beyond its glacial level by ~6 sverdrup (Fig. 1D). This BA overshoot, which generates a strong warming over the Nordic Sea region (Fig. 1, E to H, and fig. S6C), is caused by a natural overshoot of the AMOC at the end of MWF, which is further enhanced by the long duration of the MWF and the large $CO_2$ rise

from H1 to the BA. The overshoot appears to be caused by convective instability in the Nordic Sea on a background of deep-ocean warming (Fig. 2, E and F), as well as a basin-wide advective adjustment of salinity (SOM Text 4).

In contrast to the robust warming magnitude, the warming rate of the BA in CCSM3 depends critically on the MWF scenario. A faster reduction in MWF induces a more rapid BA warming, as seen by comparing experiments DGL-A and DGL-B. The sudden termination of MWF in DGL-A produces an abrupt BA warming that strongly resembles proxy records (Fig. 1), although the model warming is somewhat slower than in these records (300 versus 150 years, fig. S4). Because the natural adjustment time of the AMOC is ~300 years (as seen in experiment DGL-A), a similar abrupt BA warming can be induced in this model as long as there exists a substantial reduction of MWF several centuries before the Bølling onset. Because the two MWF scenarios DGL-A and DGL-B represent the end members, a more realistic MWF scenario that lies in between may also induce a rather rapid BA warming in CCSM3 (42). We conclude that the critical factor for producing the abruptness of the BA warming in CCSM3 is that the MWF to the North Atlantic keeps the AMOC near its off-state to within centuries of the Bølling onset, while other details of the meltwater history may not be essential.

We find that CCSM3 is able to simulate an abrupt BA warming as a transient response to a
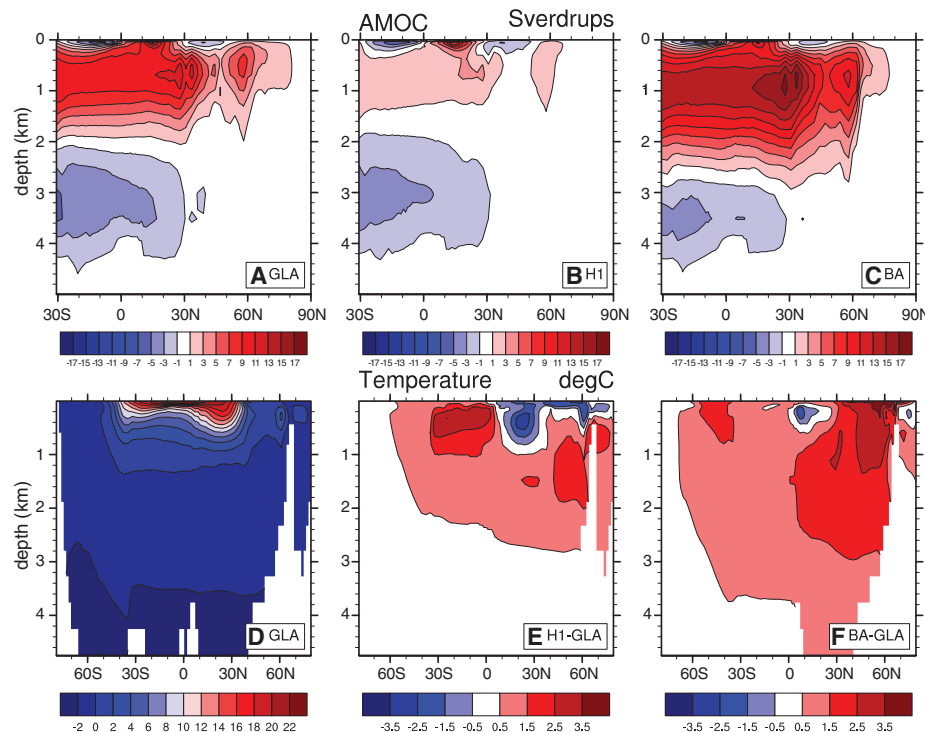


**Fig. 2.** AMOC streamfunction (in sverdrup) and Atlantic zonal mean temperature and temperature changes in experiment DGL-B (in °C). AMOC at (**A**) GLA, (**B**) H1, and (**C**) BA. Temperature at (**D**) GLA and temperature changes from the glacial state for (**E**) H1-GLA and (**F**) BA-GLA. The AMOC collapses in H1 and overshoots beyond the glacial state at BA. The subsurface ocean warms throughout the Atlantic, whereas the SST exhibits a bipolar seesaw response. (Each state is defined in Fig. 1D).

sudden termination of the North Atlantic MWF. The amplitude of the simulated BA is comparable to the amplitude reconstructed from paleoclimate proxies, notably in the North Atlantic and Antarctic regions. Our simulation suggests that the large BA warming is caused by the superposition of climate responses to increased atmospheric $CO_2$, the recovery of the AMOC from H1, and an AMOC overshoot. It remains uncertain if CCSM3 is successful in simulating the abruptness of the BA warming inferred in proxy records. CCSM3 can produce the abruptness of the BA onset only if the MWF in the North Atlantic terminates within centuries before the BA. This is in contrast to previous work in simplified models, which exhibit a substantial AMOC hysteresis (39) such that an abrupt warming can be induced by a gradual change in MWF. The behavior of CCSM3, however, is typical of the current generation of CGCMs without flux adjustment, which show little sign of substantial hysteresis (30, 43). Therefore, our results suggest that the current generation of CGCMs, like CCSM3, may not be able to induce an abrupt onset of BA warming under a gradual forcing. Is the current generation of CGCMs deficient in generating the abruptness of climate changes (44)? Is the AMOC hysteresis a fundamental feature of the real-world AMOC as suggested in intermediate models, or not essential as suggested in current CGCMs? Current observations are insufficient to address these questions unambiguously (SOM Text 5). We suggest that the critical observational evidence needed to clarify these fundamental issues is an accurate reconstruction of the rate of MWF before the BA.



**Fig. 3.** Surface air temperature changes (in °C) in experiment DGL-B. (**A**) H1 temperature anomaly from the glacial state, (**B**) BA temperature anomaly from H1, and (**C**) BA temperature anomaly from GLA. The H1 temperature response exhibits a bipolar seesaw; the BA warming is dominated by a maximum warming at northern high latitude (B) over H1, but exhibits a more symmetric warming (C) over the glacial state. (Each state is defined in Fig. 1D.)

**References and Notes**
1. W. S. Broecker, *Paleoceanography* **13**, 119 (1998).
2. P. U. Clark, N. G. Pisias, T. F. Stocker, A. J. Weaver, *Nature* **415**, 863 (2002).
3. P. M. Grootes et al., *Nature* **366**, 552 (1993).
4. R. B. Alley, P. U. Clark, *Annu. Rev. Earth Planet. Sci.* **27**, 149 (1999).
5. J. F. McManus, R. Francois, J.-M. Gherardi, L. Keigwin, S. Brown-Leger, *Nature* **428**, 834 (2004).
6. E. Boyle, L. Keigwin, *Nature* **330**, 35 (1987).
7. J. P. Severinghaus, E. J. Brook, *Science* **286**, 930 (1999).
8. C. Waelbroeck et al., *Paleoceanography* **13**, 272 (1998).
9. E. Bard, F. Rostek, J.-L. Turon, S. Gandreau, *Science* **289**, 1321 (2000).
10. D. W. Lea, D. K. Pak, L. C. Peterson, K. A. Hughen, *Science* **301**, 1361 (2003).
11. L. C. Peterson, G. H. Haug, K. A. Hughen, U. Rohl, *Science* **290**, 1947 (2000).
12. In these simplified models, however, the transient simulations have not been carried out in a synchronously coupled mode under the complete climate forcing (notably the meltwater forcing), (45, 46).
13. A. Ganopolski, S. Rahmstorf, *Nature* **409**, 153 (2001).
14. G. Knorr, G. Lohmann, *Geochem. Geophys. Geosyst.* **8**, Q12006 (2007).
15. A. J. Weaver, O. A. Saenko, P. U. Clark, J. X. Mitrovica, *Science* **299**, 1709 (2003).
16. G. Knorr, G. Lohmann, *Nature* **424**, 532 (2003).
17. The nature of the AMOC resumption is complex from the perspective of a dynamic system (47) and is beyond the scope of this study.
18. S. Rahmstorf, *Nature* **378**, 145 (1995).
19. S. G. Yeager, C. A. Shields, W. G. Large, J. J. Hack, *J. Clim.* **19**, 2545 (2006).
20. B. L. Otto-Bliesner et al., *J. Clim.* **19**, 2526 (2006).
21. A. Berger, *J. Atmos. Sci.* **35**, 2362 (1978).
22. F. Joos, R. Spahni, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1425 (2008).
23. W. R. Peltier, *Annu. Rev. Earth Planet. Sci.* **32**, 111 (2004).
24. For technical reasons, river run-off routing is kept as at the present.
25. B. L. Otto-Bliesner et al., *Geophys. Res. Lett.* **34**, L12706 (2007).
26. J. Lynch-Stieglitz et al., *Science* **316**, 66 (2007).
27. P. U. Clark, A. C. Mix, *Quat. Sci. Rev.* **21**, 1 (2002).
28. T. Crowley, *Paleoceanography* **7**, 489 (1992).
29. T. F. Stocker, *Science* **282**, 61 (1998).
30. R. Stouffer et al., *J. Clim.* **19**, 1365 (2006).
31. S. W. Hostetler, P. U. Clark, P. J. Bartlein, A. C. Mix, N. G. Pisias, *J. Geophys. Res.* **104**, 3947 (1999).
32. S.-I. Shin, Z. Liu, B. Otto-Bliesner, J. Kutzbach, S. J. Vavrus, *Geophys. Res. Lett.* **30**, 1096 (2003).
33. A. Timmermann, O. Timm, L. Stott, L. Menviel, *J. Clim.* **22**, 1626 (2009).
34. K. Cuffey, G. Clow, *J. Geophys. Res.* **102**, 26383 (1997).
35. J. Jouzel et al., *Science* **317**, 793 (2007).
36. R. Zhang, T. Delworth, *J. Clim.* **18**, 1853 (2005).
37. M. K. Davey et al., *Clim. Dyn.* **18**, 403 (2002).

38. This subsurface warming is qualitatively similar to that of previous experiments with North Atlantic MWF, especially in the tropical region and Southern Hemisphere [e.g., (30, 48)]; it seems also to be consistent with benthic $\delta^{18}O$ inferred temperature changes in the tropical Atlantic (4), western subtropical North Atlantic (4), and Nordic Sea (50), as well as Mg/Ca-based benthic temperatures from the eastern subtropical North Atlantic (51).

39. S. Rahmstorf et al., Geophys. Res. Lett. 32, L23605 (2005).

40. The MWF varies slowly during this period at a rate of ~0.1 sverdrup/ky in DGL-B, comparable to that used in previous standard hysteresis tests (18, 39).

41. About a third of the model BA warming appears to be associated with a nonlinear convective response in the Nordic Sea. As seen in experiment DGL-B, under a gradual forcing, the AMOC and Greenland temperature, after a long period of gradual change, increase abruptly at the end stage (14.6 ka) in 100 years (6 sverdrup and 6°C), nearly identical to the last 100 years of warming in experiment DGL-A (Fig. 1, D and E, and fig. S4). This abrupt warming is induced by regional convective instability in the Nordic Sea (18) on a background deep-ocean warming (52) and is enhanced by the northward heat transport associated with the enhanced AMOC.

42. With a slower reduction in MWF, the AMOC resumption and attendant BA warming in CCSM3 usually occur faster than the rate at which the MWF is reduced, partly due to

the convective instability process in the Nordic Sea (41). In a deglaciation sensitivity experiment similar to DGL-B, but with a faster termination time of 2000 years, the BA warming takes only 500 years, substantially faster than the MWF itself (not shown).

43. Except for an earlier generation of CGCM with flux adjustments (53, 54), all published results from CGCMs are similar to those of CCSM3 in showing that the AMOC recovers its strength after the termination of the freshwater pulse. Because these CGCM hosing experiments are usually short (<1000 years), they are not strict tests for the hysteresis of AMOC (18). Nevertheless, these models have shown little sign of substantial hysteresis.

44. A diagnosis of nine coupled climate models, including both CGCMs and intermediate models, suggests a positive bias in the freshwater transport by the AMOC in the South Atlantic, which may imply a bias toward a more stable AMOC and therefore a lack of multiple equilibria and in turn substantial hysteresis in these climate models (55). Further study is needed, however, to understand the stability of the AMOC in coupled climate models, especially in state-of-art CGCMs.

45. D. J. Lunt, M. S. Williamson, P. J. Valdes, T. M. Lenton, Clim. Past Discuss. 2, 267 (2006).

46. O. Timm, A. Timmermann, J. Clim. 20, 4377 (2007).

47. J. Abshagen, A. Timmermann, J. Phys. Oceanogr. 34, 2756 (2004).

48. S. Manabe, R. Stouffer, Paleoceanography 12, 321 (1997).

49. C. Ruhlemann et al., Paleoceanography 19, PA1025 (2004).

50. T. L. Rasmussen, E. Thomsen, Palaeogeogr. Palaeoclimatol. Palaeoecol. 210, 101 (2004).

51. L. Skinner, N. Shackleton, Quat. Sci. Rev. 25, 3312 (2006).

52. J. Mignot, A. Ganopolski, A. Levermann, J. Clim. 20, 4884 (2007).

53. S. Manabe, R. Stouffer, J. Clim. 1, 841 (1988).

54. J. Yin, R. Stouffer, J. Clim. 20, 4293 (2007).

55. S. Weber et al., Clim. Past Discuss. 3, 51 (2007).

56. We thank A. Ganoposki, J. Marotzke, and A. Timmermann for helpful discussions and two reviewers for comments. This research was supported mainly by the Paleoclimate Program of NSF, NCAR, and Chinese NSF (NSFC40875058). The computing is supported by the U.S. Department of Energy INCITE program and Abrupt Climate Change Program. This paper is CCR contribution No. 980.

# Undulatory Swimming in Sand: Subsurface Locomotion of the Sandfish Lizard

Ryan D. Maladen,[1] Yang Ding,[2] Chen Li,[2] Daniel I. Goldman[1,2]*

The desert-dwelling sandfish (*Scincus scincus*) moves within dry sand, a material that displays solid and fluidlike behavior. High-speed x-ray imaging shows that below the surface, the lizard no longer uses limbs for propulsion but generates thrust to overcome drag by propagating an undulatory traveling wave down the body. Although viscous hydrodynamics can predict swimming speed in fluids such as water, an equivalent theory for granular drag is not available. To predict sandfish swimming speed, we developed an empirical model by measuring granular drag force on a small cylinder oriented at different angles relative to the displacement direction and summing these forces over the animal movement profile. The agreement between model and experiment implies that the noninertial swimming occurs in a frictional fluid.

T he locomotion of organisms (1, 2), whether by running, flying, swimming, or crawling, is the result of multiple-degree-of-freedom nervous and musculoskeletal systems interacting with an environment that often flows and deforms in response to movement. Nearly all experiments and models of terrestrial locomotion have been developed for running and walking on rigid, flat, no-slip frictional substrates for which the complication of substrate flow is not considered. In contrast, complexity in interaction with the environment in aquatic and aerial locomotion (swimming and flying) is well recognized (3). Determining mechanisms for propulsion or lift in these media is

[1]Interdisciplinary Bioengineering Program, Georgia Institute of Technology, Atlanta, GA 30332, USA. [2]School of Physics, Georgia Institute of Technology, Atlanta, GA 30332, USA.

*To whom correspondence should be addressed. E-mail: daniel.goldman@physics.gatech.edu

always possible in principle because the rules of interaction with fluids are worked out: They require solving Navier-Stokes hydrodynamics in the presence of moving boundary conditions. A major challenge in biology is to understand the locomotion of organisms that walk, crawl, or burrow on or within terrestrial substrates like sand (4), soil (5), and muddy sediments (6) that display both solid and fluidlike behavior. In such materials, validated theories such as the Navier-Stokes equations for fluids do not exist, and visualization techniques [such as particle image velocimetry in fluids (7)] are nearly nonexistent. Understanding of the mechanics of subsurface movement has ecological importance and could reveal how the actions of small burrowing organisms can transform entire landscapes (8).

Animal burrowing and movement within granular media is relevant to desert organisms like scorpions, snakes, and lizards that move within sand to escape heat and predators and hunt

for prey (9, 10). Desert sand [which covers 6 to 10% of land surface (11)] is an example of a granular material, a collection of dissipative particles that interact through contact forces and in bulk can display solid and fluidlike features (12) when disturbed. A key parameter that controls the response of granular media to intrusion is the volume fraction φ, the ratio of material volume to total occupied volume. In dry granular media in natural environments, φ depends on the history of the sand (for example, perturbations by wind or animal burying and digging), and can vary between 0.57 and 0.64 (13). The response of granular media to intrusion depends on φ: Closely packed material at high φ must expand to flow, whereas loosely packed material at low φ consolidates (14). The effects of φ on drag are largely unexplored, although we have recently found that vertical penetration resistance doubles as φ increases by just 0.08 (15).

To investigate how rheological features of the material influence the locomotor mode and performance of an organism moving within sand, we used high-speed x-ray imaging to study a small (~10 cm) desert-dwelling lizard, the sandfish, that inhabits the Saharan desert of Africa and moves within granular media of different φ. The sandfish's above-ground burial process has been described (9), and it is hypothesized that its counter-sunk lower jaw and smooth scales with low friction and low wear properties (16) aid swimming and digging. However, little is known about how the animal moves subsurface. Although it has been hypothesized that body motion plays an important role in thrust production (9, 17) in sand-dwelling lizards, a recent study using nuclear magnetic resonance (NMR) to visualize subsurface motion proposed that the sandfish used its limbs in a paddling motion along with undulations on its body to generate thrust subsurface (18). However, the observa-

tional quality and number of trials were limited. We show through detailed biological experiment and development of a model of the locomotion that subsurface sandfish motion is not driven by limbs but by an undulatory gait without use of limbs. Thrust and drag result from frictional grain flow, which leads to features of the swimming that are intermediate to movement in fluids and on solid surfaces, and surprisingly are independent of φ.

To perform controlled studies of the burial and swimming of the sandfish (11 animals; mass = 16.2 ± 4 g, body length $L$ = 8.3 ± 3.3 cm, defined as the length from snout tip to the base of the tail), we used a fluidized bed (fig. S1) (*19*) to prepare a 10-cm-deep container of 0.27 ±

0.04 mm glass beads into naturally occurring loosely packed and closely packed volume fractions φ = 0.58 and φ = 0.62, respectively. The sandfish took 672 ± 239 ms to complete its entire burial process, independent of volume fraction (Student's $t$ test, $P$ > 0.05) (movie S1). Once below the surface, the dorsal x-ray video imaging (movies S2 and S3) revealed that the animal no longer used limbs for propulsion. Instead, the animal placed its limbs against its sides and executed an undulatory motion (Fig. 1, E and F) with large-amplitude axial oscillation, using its body to propel itself at speeds of up to ~10 cm/s. We observed similar limbless undulatory motion in beads with mean diameters ranging from 0.1 mm to 3 mm. To quantify the lack of limb use, we measured the distance between pairs of limbs (fore and hind) (Fig. 1G). As the animal ran across the surface, this inter-limb distance oscillated around 2.75 ± 0.59 cm, indicating the sprawl of the limbs during the diagonal gait (*1*, *9*). As it began to bury, the inter-limb distance decreased significantly (Student's $t$ test, $P$ < 0.01) until it became a constant (1.72 ± 0.53 cm) close to body width (1.23 ± 0.51 cm) because limbs were placed close to the side of the body. Like in (*18*), side-view x-ray imaging of the sandfish subsurface showed that the animal undulates in a plane at ~ 22.2 ± 3.7° with respect to the horizontal and buries consistently to a depth of 2.1 ± 0.5 cm (surface to top of mid-point of the back), independent of volume fraction (Student's $t$ test, $P$ > 0.05).

We characterized the undulatory kinematics by digitizing the midline from the dorsal x-ray images (Figs. 1F and 2A). The undulatory motion fit well ($r^2$ > 0.85, $P$ < 0.01) the form of a posteriorly traveling single-period sinusoidal wave (Fig. 2A) such that $y = A\sin 2\pi(x + v_w t)/\lambda$, with $y$ the displacement away from the midline of a straight animal, $A$ the amplitude, $\lambda$ the wave-length, $f$ the wave frequency, and $v_w = f\lambda$, the wave speed. From the fit, we deduced these parameters as well as the forward swimming speed of the center of mass $v_x$ (which was not equal to the wave speed; see discussion below). During a swim, there was no change in either $A$ ($P$ = 0.12) or $\lambda$ ($P$ = 0.66), and neither $A$ nor $\lambda$ varied significantly for the different φ ($P$ > 0.05 comparing each treatment by using a repeated measures one-way analysis of variance) (Fig. 2B). The large-amplitude undulations over the entire body are unlike the kinematics of other undulatory swimming organisms on the sandfish-sized scale [such as eels (*20*)], which tend to increase in amplitude posteriorly. The undulations do resemble those of eel movement on land (*20*) as well as small (<1 mm) swimmers (for example, nematodes) in fluids at low Reynolds (*Re*) number, in which the inertia of the material surrounding the organism plays no role (*21*, *22*).

The sandfish increased swimming speed by increasing wave frequency $f$ (Fig. 2C) (Student's $t$ test, $P$ < 0.01), and the slope of the $v_x$ versus $f$ relationship was independent of vol-
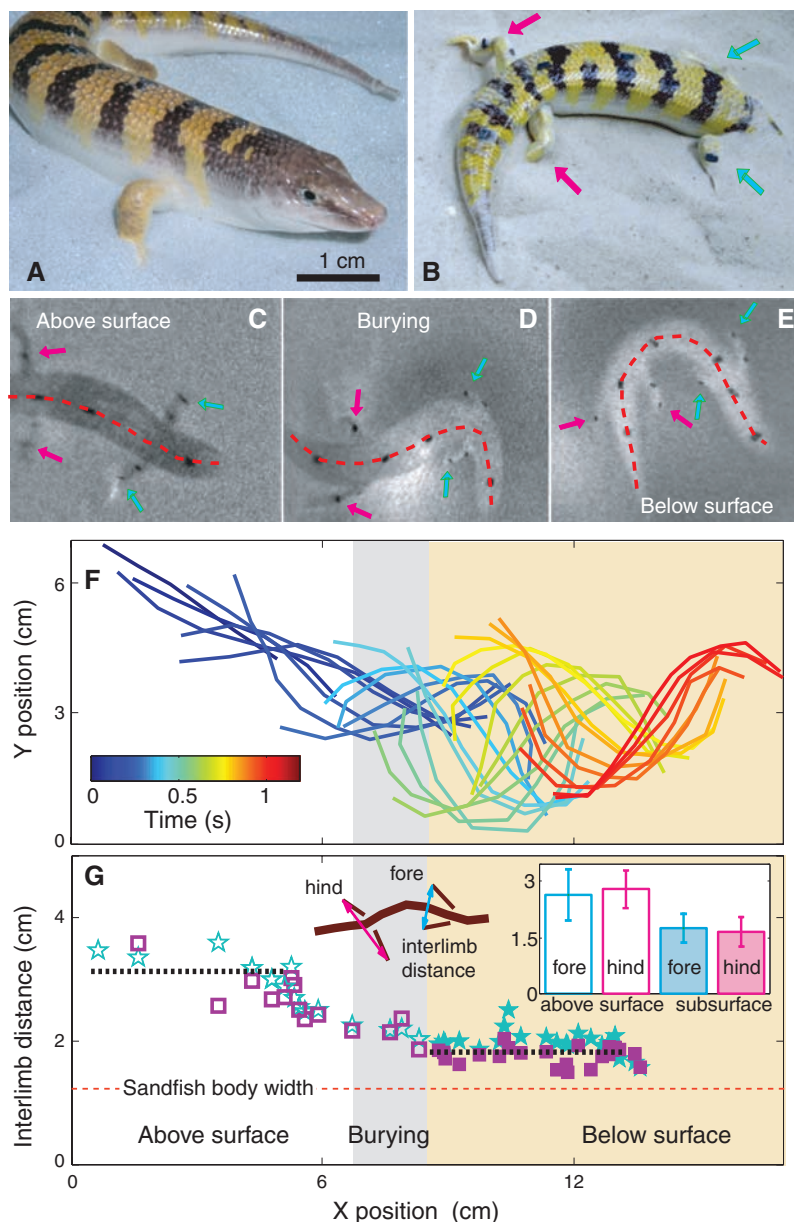


**Fig. 1.** Burial and undulatory subsurface swimming of the (**A**) sandfish lizard. (**B**) The burying sandfish with opaque makers bonded to its fore (cyan) and hind (purple) limbs and midline, just before it dives into the media. (**C**) Representative x-ray images of the sandfish as it moves with a diagonal gait on the surface, (**D**) as it buries into the media with high-amplitude oscillations and use of limbs, and (**E**) moving subsurface with its limbs at its sides. (**F**) Curves show the tracked midline of the sandfish as the animal runs on the surface using a diagonal gait (*1*, *9*) with little back bending (unshaded region), buries with significant back bending (gray region from when snout first enters material to when fully submerged), and then swims within the media with large-amplitude back bending (brown shading). The curve color (color bar, inset) indicates time progress of the motion. (**G**) Interlimb distance of the fore (blue star) and hind (purple square) limbs versus the position of the center of mass. The dashed lines represent the average interlimb distance (10 animals) above and below the surface. (Inset) Bar graph shows the interlimb distance (mean ± SD) for the fore and hind limbs both above and below the surface. *n* = 46 trials were recorded.

ume fraction (Student's *t* test, $P > 0.05$). We found, however, that $v_x$ is not simply the product of $f$ and $\lambda$, but instead the slope is reduced by a constant factor $\eta$ so that $v_x = \eta f$ (Fig. 2C) with $\eta = 0.54 \pm 0.12$ and $\eta = 0.49 \pm 0.09$ for loosely and closely packed material, respectively. Because $v_w = \lambda f$, this implies that the average forward speed is a fraction of the wave speed. Thus, the animal is not moving in a tube because successive undulations do not trace a continuous path; this can be seen in Fig. 1F. Instead, tracer particles placed in the bulk reveal that there is backward displacement of material as the animal moves forward. Slipping while progressing is common to undulatory swimmers in deformable media across length scales (for example, eels and spermatozoa in fluids) (*1*) and is characterized by the wave efficiency ($\eta$) (*21*), defined as the ratio between the forward speed of the animal ($v_x$) and the velocity of the wave traveling down its body ($v_w$): $\eta = v_x/v_w$. Because $\lambda$ for the sandfish is independent of $f$ in our experiments, $\eta$ is the slope of the $v_x/\lambda$ versus $f$ curve and is shown in Fig. 2C. Typical wave efficiencies of organisms in moving in fluids at low *Re* (such as nematodes) are 0.25 (*23*), nearly a factor of two smaller than the sandfish, whereas organisms undulating (creeping) along the solid surface of an agar-air interface can have $\eta$ close to unity [$\eta \sim 0.8 - 0.9$ (*21*)] because deformations in the surface allow movement that is effectively in a tube. $\eta$ for the sandfish in granular media is intermediate to that for fluids and solids and is independent of $\phi$, even though materials with different $\phi$ have penetration resistances that can vary by a factor of 2 (*15*).

Prediction of $\eta$ for undulatory movement in fluids and solids has been accomplished by use of models of interaction of the organism and the environment: In low *Re* Newtonian fluids, forces are purely viscous (Stoke's law) such that force is proportional to instantaneous velocity (*22, 23*), and in undulatory crawling on solids force is determined with static/dynamic friction coefficients (*22*). Although the equivalent force laws are not available in granular media, features of the sandfish motion allow us to hypothesize about the medium in which it swims: (i) It is well known that for slow enough disturbances, the drag in granular systems is dominated by friction between flowing grains, and thus this makes drag independent of velocity (*14, 24–26*). (ii) At the observed body speeds and accelerations during undulatory motion, inertial forces due to accelerating material are small as compared with frictional drag forces ($ma/F \approx 0.2$, where $F \approx 2$ N is estimated for a small cylinder dragged in sand and $ma \approx 0.4$ N is found from the product of the acceleration of the sandfish body and a mass of sand surrounding it), and thus inertia can be neglected. (iii) Because grain interactions are dissipative, disturbances typically dampen in short times, and thus changes in the static configuration dominate the drag force.

Therefore, we hypothesize that the animal swims in a so-called "frictional fluid" (*27*), such that flowing grain-grain and grain-animal friction

determine the thrust and drag forces. Such a frictional fluid superficially resembles low *Re* swimming in that there is no inertia, but the mechanism for drag is frictional, rather than viscous. Movement in a frictional fluid differs from noninertial crawling on a surface in which anisotropic friction [originating from, for example, deforming the surface and pushing off as does nematode on agar (*21*) or using overlapping belly scales or changes in weight distribution as does a snake (*28*)] generates net propulsive force without net flow of material.

To test this hypothesis, we developed a model inspired by similar local force, noninertial models [typically called Resistive Force Theory (RFT) (*23*)] such that the body of the organism is partitioned into elements, each of which generates thrust and experiences drag when moving through a media and whose forces are not influenced by force fields from other segments. If $F_N$ and $F_L$ are the normal and lateral forces (perpendicular and parallel to the element respectively) acting on the sides of the sandfish body, net forward force on the element can be written as $F_x = F_N \sin\theta - F_L \cos\theta$, where $\theta$ is the angle between the forward direction of the animal and the orientation of an element of the organism (Fig. 3A). The forces for each element are integrated over the length of the body (snout to tail tip) and for a full period of a traveling wave to obtain average swimming speed. Unlike in fluids, we have no validated theory to calculate the force on an intruder as a function of its angle relative to displacement or

for varying $\phi$. Therefore, we used a combination of measurements (*19*) of the granular thrust and drag forces on a stainless steel cylindrical rod [with grain-surface friction coefficient similar to the sandfish (*16*)], insights from Fig. 3B, and an empirical fitting function that incorporates drag on the sidewalls and end-faces of the cylinder (equation S8) so as to approximate $F_N$ and $F_L$ for elements of the sandfish body as:

$$F_N = 2lr(C_S \sin\beta_0 + C_F \sin\psi)$$
$$F_L = 2lrC_F \cos\psi \qquad (1)$$

where $\tan\beta_0 = \cot\gamma_0 \sin\psi$, $\psi$ is the angle of the segment with respect to its local velocity vector, $l$ and $r$ are the length and radius of the segment (assumed cylindrical), and the constants $C_S$, $C_F$, and $\gamma_0$ characterize the material response to drag. Plots of $F_N$ and $F_L$ as a function of $\psi$ are shown in Fig. 3, C and D. Equation 1 represents the data well ($R^2 = 0.97$) (solid lines shown in Fig. 3, C and D), and the fit coefficients are given in table S1.

$F_L$ is well approximated by the cosine function, similar to viscous fluids or frictional material, indicating that its source is the resolved component of friction between the grains and the body. The form of $F_N$ is not well approximated by the corresponding sine component (Fig. 3C, dashed gray lines) but rather increases more rapidly for $\psi < 30°$ and increases more slowly for $\psi > 30°$. We hypothesize that above this angle [close to the angle at which internal slip planes or shear bands form and move (*14*)], the force

**Fig. 2.** Kinematics of the undulatory sandfish motion. (**A**) Traveling wave moving down the body of the sandfish opposite to the direction of the sandfish forward motion (sampled every 0.04 s). For each time instant, the instantaneous lateral displacement of a tracked section of the sandfish is represented in color. The black curves represent the tracked midline (for example, Figure 1E, snout tip to tail tip) of the sandfish. (**B**) The amplitude $A$ and wavelength $\lambda$ of the sinusoidal fit to the tracked midline of the sandfish (inset), for loosely and closely packed media.



(**C**) Red circles (closely packed, $n = 24$ trials), and blue triangles (loosely packed, $n = 22$ trials) show dependence of swimming velocity on wave frequency. The slope of green line is the approximate wave efficiency $\eta \approx 0.5$ of animal (Fig. 4A) and the dashed black line represents $\eta = 1$. The horizontal and vertical blue and red bars (mean $\pm$ 2 SD) represent the range of frequencies and forward velocities (normalized by $\lambda$) accessed in loosely packed ($2.10 \pm 0.75$ Hz, $0.98 \pm 0.27$ per second) and closely packed ($2.63 \pm 1.13$ Hz, $1.28 \pm 0.37$ per second), respectively. The average frequency and average speed are greater in closely packed than in loosely packed material (Student's *t* test, $P < 0.01$). $n = 46$ trials were recorded.
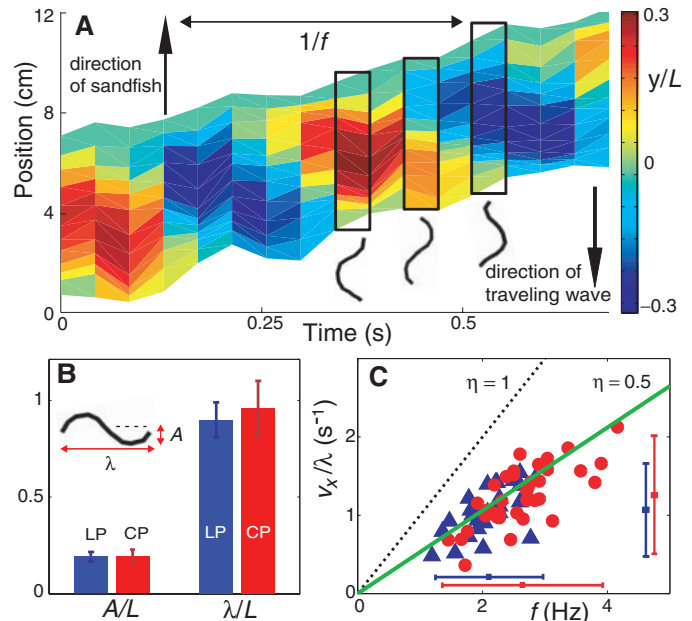
increases slowly because of the formation of a solid region of grains that moves along with the rod (14, 26). Preliminary numerical simulation indicates that the size of the solid region grows slowly with ψ for ψ > 20°. The fitting function (Eq. 1) incorporates these effects in $F_N$ and $F_L$ through terms proportional to the angular component of $C_S$ (which is determined primarily by grain-object frictional forces) and an extra term in $F_N$ that increases rapidly and then saturates for ψ > 20°. In this term, the coefficient $C_S$ is set by grain-grain frictional force and $\gamma_0$, which is related to the internal slip angle. $C_F \approx C_S/2$, indicating that grain-grain friction is larger than grain-surface friction.

To determine the wave efficiency (and thus swimming speed for a given frequency), we balanced thrust and drag over a cycle (net force, $\overline{F}_x = 0$ in equation S7, assuming constant cycle average speed $v_x$). Because $v_x$ and $v_w$ appear only as a ratio (from the kinematics in equation S2) in the argument of the integral in Equation S7, and do not appear in the $F_N$ and $F_L$ terms (fig. S2B), the force balance predicts η. Numerical integration of the model over a measured organism sinusoid from Fig. 2 with constant cylindrical cross-section predicts $0.4 \leq \eta \leq 0.65$ independent of φ (Fig. 4A), which is in accord with experimental observation; the bounds on η are set by assuming the head drag is zero or that of a flat plate because measurements for the drag force on an object with the exact morphology of the shovel-shaped head are unavailable. The model predicts a maximum in forward progress per cycle when $A/\lambda \approx 0.2$, and we find that the measured kinematics fall close to this speed (Fig. 4B). The maximum is due to a competition between increased η (less slipping) from increased $A$ and decreased λ because for an inextensible animal λ decreases as $A$ increases. An analytic low-amplitude approximation (equations S9 to S14) demonstrates the relevant scaling of η.

It is remarkable that η does not change significantly for different φ, especially because drag forces between closely and loosely packed material differ by ≈ 80%. The model indicates that η is essentially constant because $F_N$ and $F_L$ scale by the same ratio of 2.1 between closely and loosely packed material (Fig. 3, C and D, insets), and it is this ratio that determines η (equation S14). The model implies that η is greater than in low-*Re* fluids because of the greater thrust to drag force ratio in the frictional granular media.

The agreement between experiment and model indicates that the assumptions of velocity independence, local interaction, and the use of the average $F_N$ and $F_L$ for all phases of the motion are good approximations at these slow swimming speeds. Although the assumption that all body and head segments encounter material with the same φ is probably not true (because the body encounters grains disturbed by the head and other body segments), because $F_N$ and $F_L$ scale in the same ratio in all φ and this ratio determines η, wave efficiency should be independent of position along the body. The model assumptions break down in the approximation that the instantaneous force (Fig. 3B) can be replaced by the average force. This is a good approximation except for $F_N$ in closely packed material (Fig. 3B, red curves); enhancement of $F_N$ because of peak stress significantly greater than the mean could explain why, contrary to expectation, the range of $f$ (and $v_x$) that the animal accessed was greater in closely packed material and the maximal swimming speed the animal was able to access was ≈50% larger in closely packed material (Fig. 2C).

In conclusion, we used high-speed x-ray imaging to show that the sandfish lizard propels itself within granular media using a large-amplitude traveling-wave oscillation of its body without using its limbs and that the wave efficiency of this motion is independent of the volume fraction of the medium. Using a model that balances thrust and drag with measured granular drag laws, we can predict the wave efficiency and optimal kinematics. We demonstrated that the animal swims in a medium with properties intermediate to those of low *Re* viscous fluids and frictional solid surfaces and that are unique to dense granular beds, namely velocity-independent forces and enhanced normal-drag forces. Our results demonstrate that burrowing and swimming in complex media can have similar complexity to



**Fig. 3.** Drag measurements to model undulatory locomotion: (**A**) Diagram of forces $F_N$ (perpendicular) and $F_L$ (parallel) on a given element of a sinusoidal traveling wave. (**B**) Representative total drag force profiles (including sidewalls and end-faces) versus time for a stainless-steel cylinder (diameter = 1.58 cm and length = 4.00 cm) oriented at an angle ψ = 60° to the direction of its forward velocity (10 cm/s). The solid lines correspond to $F_N$ and the dashed to $F_L$. The red and blue colors correspond to closely and loosely packed media, respectively. The label $A$ represents displacement associated with the average amplitude of oscillation of the sandfish. (**C**) Average $F_N$ and (**D**) $F_L$ on the cylinder (removing force contribution from the end-faces); red circles and blue triangles correspond to closely and loosely packed preparations, respectively. In (C) and (D), the average forces corresponding to the force profiles in (B) are indicated with the respective rectangles. Both insets show the loosely packed force data scaled by 2.1. Averages of $F_N$ and $F_L$ are calculated over the shaded region in (B). Solid lines are model fits described by Eq. 1. For comparison, dashed gray lines correspond to $F_N$ and $F_L$ calculated for a fluid (water) by choosing a velocity that fits $F_L$ versus ψ.



**Fig. 4.** Comparison of experimental results and model predictions; blue denotes loosely packed and red denotes closely packed material. (**A**) Animal wave efficiencies from experiment are represented by solid bars (loosely packed average η = 0.54 ± 0.12; closely packed average η = 0.49 ± 0.09). Shaded bars correspond to model predictions assuming head drag of a flat disk (FH) and no head drag (NH). (**B**) Predicted swimming speed (in body-length per cycle) versus amplitude normalized by wavelength. The dashed lines are model predictions with no-head (short-dashes) and flat-head (long-dashes) for loosely packed and closely packed material. Squares denote the measured average animal swimming speed for loosely packed and closely packed material. The range of frequency and speed accessed is given by the extent of the solid lines. $n$ = 46 trials were recorded.

movement in air or water, and that organisms can exploit the solid and fluidlike properties of these media to move effectively within them.

### References and Notes

1. R. M. Alexander, *Principles of Animal Locomotion* (Princeton Univ. Press, Princeton, NJ, 2003).
2. M. H. Dickinson et al., *Science* **288**, 100 (2000).
3. S. Vogel, *Life in Moving Fluids* (Princeton Univ. Press, Princeton, NJ, 1994).
4. E. R. Trueman, *J. Exp. Biol.* **53**, 701 (1970).
5. C. R. Darwin, *The Formation of Vegetable Mould, Through the Action of Worms: With Observations on Their Habits* (John Murray, London, 1881).
6. K. Dorgan, P. Jumars, B. Johnson, B. Boudreau, E. Landis, *Nature* **433**, 475 (2005).
7. G. Lauder, E. Drucker, *Physiology (Bethesda)* **17**, 235 (2002).
8. F. Meysman, J. Middelburg, C. Heip, *Trends Ecol. Evol.* **21**, 688 (2006).
9. E. Arnold, *J. Zool.* **235**, 351 (1995).
10. C. White, *Aust. J. Zool.* **49**, 663 (2001).
11. E. Ezcurra, *Global Deserts Outlook* (United Nations Environmental Programme, Nairobi, Kenya, 2006).
12. H. M. Jaeger, S. R. Nagel, R. P. Behringer, *Phys. Today* **49**, 32 (1996).
13. W. W. Dickinson, J. D. Ward, *J. Sediment. Res. Sect. A* **64**, 226 (1994).
14. R. Nedderman, *Statics and Kinematics of Granular Materials* (Cambridge Univ. Press, New York, 1992).
15. C. Li, P. B. Umbanhowar, H. Komsuoglu, D. E. Koditschek, D. I. Goldman, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3029 (2009).
16. W. Baumgartner et al., *J. Bionic Eng.* **4**, 1 (2007).
17. C. Gans, *Integr. Comp. Biol.* **15**, 455 (1975).
18. W. Baumgartner et al., *PLoS One* **3**, (e3309), 1 (2008).
19. Materials and methods are available as supporting material on *Science* Online.
20. A. A. Biewener, G. B. Gillis, *J. Exp. Biol.* **202**, 3387 (1999).
21. J. Gray, H. Lissman, *J. Exp. Biol.* **41**, 135 (1964).
22. J. Korta, D. Clark, C. Gabel, L. Mahadevan, A. Samuel, *J. Exp. Biol.* **210**, 2383 (2007).
23. J. Gray, G. J. Hancock, *J. Exp. Biol.* **32**, 802 (1955).
24. I. Albert et al., *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **64**, 61303 (2001).
25. R. Albert, M. Pfeifer, A. Barabási, P. Schiffer, *Phys. Rev. Lett.* **82**, 205 (1999).
26. K. Wieghardt, *Annu. Rev. Fluid Mech.* **7**, 89 (1975).
27. A. N. Schofield, C. P. Wroth, *Critical State Soil Mechanics* (McGraw-Hill, London, 1968).
28. B. R. Moon, C. Gans, *J. Exp. Biol.* **201**, 2669 (1998).
29. This work was supported by NSF Physics of Living Systems grant PHY-0749991 and the Burroughs Wellcome Fund. We thank R. Full and T. Papenfuss for helpful discussion. We thank S. Steinmetz and N. Gravish for experimental assistance and P. Umbanhowar for discussion of the drag model and careful reading of the manuscript.

# Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes

Adrian W. Briggs,[1]* Jeffrey M. Good,[1] Richard E. Green,[1] Johannes Krause,[1] Tomislav Maricic,[1] Udo Stenzel,[1] Carles Lalueza-Fox,[2] Pavao Rudan,[3] Dejana Brajković,[4] Željko Kućan,[3] Ivan Gušić,[3] Ralf Schmitz,[5,6] Vladimir B. Doronichev,[7] Liubov V. Golovanova,[7] Marco de la Rasilla,[8] Javier Fortea,[8] Antonio Rosas,[9] Svante Pääbo[1]

Analysis of Neandertal DNA holds great potential for investigating the population history of this group of hominins, but progress has been limited due to the rarity of samples and damaged state of the DNA. We present a method of targeted ancient DNA sequence retrieval that greatly reduces sample destruction and sequencing demands and use this method to reconstruct the complete mitochondrial DNA (mtDNA) genomes of five Neandertals from across their geographic range. We find that mtDNA genetic diversity in Neandertals that lived 38,000 to 70,000 years ago was approximately one-third of that in contemporary modern humans. Together with analyses of mtDNA protein evolution, these data suggest that the long-term effective population size of Neandertals was smaller than that of modern humans and extant great apes.

Currently, DNA sequences determined from multiple Neandertals are restricted to short fragments [120 to 360 base pairs (bp)] of the hypervariable regions (HVRs) of mitochondrial DNA (mtDNA) (*1, 2*). These data have demonstrated that Neandertal mtDNA is distinct from that of modern humans (*3*). However, collecting sequence data from the rest of the mtDNA ge-

nome has proven difficult due to numerous technological difficulties (*4*). Recently, the complete mtDNA genome sequence of a ~38,000-year-old Neandertal individual from Vindija Cave, Croatia, was determined by high-throughput shotgun sequencing from total DNA extract (*5*). A similar approach has been used to recover complete mtDNA sequences from permafrost-preserved mammoths and a human (*6, 7*). However, the amount of shotgun sequencing needed to retrieve complete mtDNA sequences is prohibitive for most ancient bone specimens due to the high fraction of environmental DNA that they contain. For example, only 0.001% of DNA sequences determined from typical well-preserved Neandertal specimens are derived from mtDNA (table S1). Thus, a simple shotgun approach would require hundreds or thousands of high-throughput pyrosequencing runs to recover a single Neandertal mitochondrial genome (table S1). Direct polymerase chain reaction (PCR) is also poorly suited for retrieving complete Neandertal mtDNA genomes, because DNA extracted from the fossils

is so fragmented that hundreds of overlapping amplicons would be necessary, either requiring highly multiplexed primer mixes that present severe difficulties for avoiding modern human contamination, or many parallel amplification reactions that consume large amounts of precious ancient DNA extracts (*8*).

We have developed a method—primer extension capture (PEC)—that directly isolates specific DNA sequences from complex libraries of highly degraded DNA (Fig. 1). PEC uses 5′-biotinylated oligonucleotide primers and a DNA polymerase to capture specific target sequences from an adaptor-ligated DNA library. It combines the high specificity of PCR primers with the numerous advantages of a library sequencing approach, including immortalization through reamplification from adaptor priming sites (*9*) (fig. S1), contamination control with project-specific barcodes (*5, 10*), access to very short fragments predominant in ancient extracts (*11*), and quantification of the number of unique ancient DNA molecules, which is necessary to identify nucleotide misincorporations (*10–12*).

We used PEC to recover the entire Neandertal mtDNA genome [supporting online material (SOM)] of five individuals from four sites across the geographic range of Neandertals (Fig. 2 and table S2). One individual (Vindija 33.25) from Vindija Cave, Croatia, is undated but was found in an older stratigraphic layer than the previously sequenced bone (*5*) (Vindija 33.16), which was dated to ~38,000 years before present (yr B.P.) (*3*). Two individuals (Feldhofer 1 and 2, the former being the Neandertal type specimen) come from Kleine Feldhofer Grotte, Neander Valley, Germany, and are dated to ~40,000 yr B.P. (*13*). One individual (Sidron 1253) comes from El Sidron Cave, Spain, and is dated to ~39,000 yr B.P. (*14*), and one (Mezmaiskaya 1) comes from Mezmaiskaya Cave, Russia, and is dated to 60,000 to 70,000 yr B.P. (*15*).

We generated between 170,330 and 521,680 sequence reads per individual on the 454 FLX platform and processed them with a mapping

[1]Max-Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany. [2]Institute of Evolutionary Biology, Consejo Superior de Investigaciones Científicas, Universitat Pompeu Fabra, 08003 Barcelona, Spain. [3]Croatian Academy of Sciences and Arts, Zrinski trg 11, HR-10000 Zagreb, Croatia. [4]Croatian Academy of Sciences and Arts, Institute for Quaternary Paleontology and Geology, Ante Kovačića 5, HR-10000 Zagreb, Croatia. [5]Landschaftsverband Rheinland Landesmuseum, D-53115 Bonn, Germany. [6]Department of Early Prehistory and Quaternary Ecology, University of Tübingen, Germany. [7]Laboratory of Prehistory, St. Petersburg, Russia. [8]Área de Prehistoria Departamento de Historia Universidad de Oviedo, Oviedo, Spain. [9]Departamento de Paleobiología, Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas, Madrid, Spain.

*To whom correspondence should be addressed. E-mail: briggs@eva.mpg.de

assembly program (SOM), taking the Vindija 33.16 mtDNA sequence (5) as a reference. Between 18.2 and 40.2% of sequences were identified as mitochondrial, representing a raw target enrichment of 3640- to 80,400-fold (table S1). Nontarget sequences were enriched in sequence motifs from the 3′ ends of PEC primers (fig. S2), suggesting the occurrence of some false priming events. All five individuals showed continuous coverage across the entire mtDNA genome with an average unique read depth of 18.0 to 56.3× (fig. S3). As reported previously for Neandertal mtDNA (5), average fragment lengths were short (51.3 to 79.3 bp, fig. S4), coverage was positively correlated with GC content (fig. S5), and bases differing from the consensus were dominated by C/G→T/A substitutions (fig. S6), consistent with deamination of cytosine (10, 11, 16) clustered toward the ends of fragments (10, 11) (fig. S7). However, the rate of C/G→T/A misincorporations varied among the fossils and was negatively correlated with average fragment length (fig. S8), likely reflecting differential hydrolytic damage suffered by DNA in different depositional environments. To determine if sequencing coverage in each assembly was sufficient to overcome misincorporations and other sequencing errors, we calculated the proportion of reads at each position that matched the assembly consensus base (fig. S9). Apart from a single ambiguous position (see fig. S10 for resolution of this position), support across the 82,865 aligned positions (five 16,565-bp mtDNA genomes) was high (mean = 98.5%, minimum 60.0%). At all 43 positions with support lower than 75%, we observed a minor base consistent with deamination of cytosine. Sites that differed from the Neandertal reference had coverage (mean = 32.0×) and support (mean = 98.7%) similar to those that matched the reference (mean coverage = 36.5×; mean support = 98.5%). None of these sites had support lower than 78% (table S3), and their locations were not correlated with PEC primer sites (all $P > 0.50$, Fisher's exact test; table S4). These data suggest that the five consensus sequences are unlikely to have been affected by sequencing or base damage errors, because such errors would tend to create sequence differences that would be poorly supported and located in regions of low coverage.

We quantified the level of modern human DNA contamination in each of our assemblies by counting Neandertal versus contaminating fragments at all positions (124 to 130 positions per individual) where the Neandertal was found to differ from a worldwide panel of 311 aligned modern human mtDNA genomes (5). From 876 to 2969 fragments overlapping these positions per Neandertal assembly, we estimated that modern human contamination ranged from 0.2 to 1.4% (table S5). At these levels, contamination is highly unlikely to affect the determination of consensus sequences. Finally, the PEC-derived Neandertal mtDNA genomes were compared to PCR-amplified HVR sequences from the same individuals (13, 17) (SOM and fig. S11). Aligned sequences were identical between the two methods for every individual except Feldhofer 1, where some previously postulated errors for PCR-derived positions (13) were confirmed as such by PEC.

Comparison among the six complete Neandertal mtDNA genomes revealed a total of 55 variable positions across 16,565 aligned nucleotides (Table 1). On average, the six Neandertal mtDNAs differ by 20.4 substitutions. We contrasted Neandertal mtDNA diversity with variation among modern humans as represented by the revised Cambridge reference sequence (18) and a previously published worldwide sample of 53 individuals (19). Variation among the Neandertals was approximately one-third of that estimated for modern humans worldwide, approximately half that of individuals from non-African populations, and similar to that of the nine Europeans in this sample (Table 1). When compared to a broader survey of 30 modern Europeans (Table 1 and SOM), Neandertals had 37% lower mtDNA diversity. Because the Neandertal sequences stem from several distinct time points spanning thousands of years, we used coalescent simulations (20, 21) to assess the magnitude of any bias introduced by the temporal sampling and found that it may cause up to a ~20% overestimate of Neandertal genetic diversity (table S6) relative to a sample drawn from a single point in time. Thus, the observation that mtDNA diversity is lower in Neandertals than in modern humans appears to be conservative with respect to sampling over time.

Estimation of phylogenetic relationships confirmed the reciprocally monophyletic status of Neandertal and modern human mtDNAs (Fig. 2). Among the six Neandertal mtDNAs, the sequence

from Mezmaiskaya is more divergent from the other five (44.4 mean pairwise differences) than the latter are among themselves (8.4 mean pairwise differences). As observed previously (5), the Neandertal branch in the phylogeny appears shorter than that of modern humans. Simulations show that this asymmetry is well within the expected range of stochastic variation given that the Neandertal mtDNA sequences are at least 38,000 years old (fig. S12).

With a Bayesian approach (22) calibrated with the fossil ages of the five dated Neandertal sequences and a human-chimpanzee divergence of ~6 million years (see SOM), we estimated the time to the most recent common Neandertal mtDNA ancestor to be ~109,800 yr B.P. [84,630 to 138,500 yr B.P., 95% highest posterior density (HPD); table S7]. This is ~80% of the mean coalescent time estimated for modern human mtDNA with the same approach (mean time to the most recent common ancestor = 136,100 yr B.P.; 94,930 to 178,700, 95% HPD). Assuming that positive selection has not influenced Neandertal mtDNA variation, a recent coalescent time despite a ~38,000-year truncation of the Neandertal lineage suggests that the effective population size of Neandertals ($N_e$) was small and probably included fewer than 3500 females (mean $N_e$ = 1476; 268 to 3510, 95% HPD). The recovery of identical mtDNA sequences from Vindija in Croatia and the Neander Valley in Germany, which are more than 850 km apart, supports the inference of a small effective population size in Neandertals. Given the temporal sampling and a constant effective size of 1500, the probability of sampling at least two identical mtDNA genomes among six is reasonably high ($P = 0.25$). In contrast,



**Fig. 1.** Primer extension capture (PEC). (i) 5′-Biotinylated oligonucleotide primers (PEC primers) are added to a 454 library [in which the A and B adaptor molecules carry a project-specific barcode (10)] and are allowed to anneal to their respective target sequences. (ii) A single *Taq* DNA polymerase extension step is performed, resulting in a double-stranded association between primer and target that includes the 5′ adaptor sequence. (iii) Excess PEC primers are removed by spin column purification, and the biotinylated primer:target duplexes are captured by streptavidin-coated magnetic beads. The beads are washed stringently above the melting temperature of the PEC primers, to ensure that templates upon which extension occurred will preferentially remain associated with the primers. (iv) Captured and washed targets are eluted from the beads, amplified with adaptor priming sites, and subjected either to a second round of extension and capture or directly to 454 emulsion PCR (full details are in the SOM).

only one pair of identical sequences was identified among the 30 Europeans, as expected given the recent population expansion of modern humans (23).
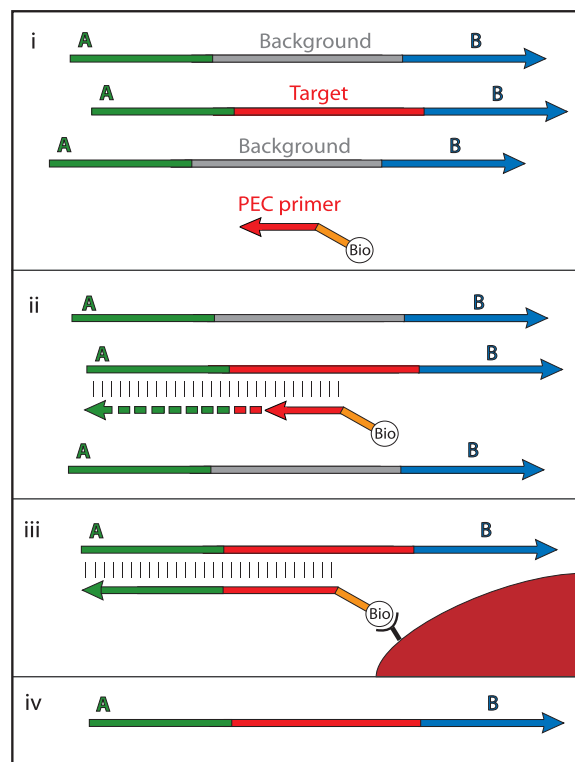
The observation that the most divergent mtDNA genome was retrieved from the oldest and easternmost individual analyzed (Mezmaiskaya 1) raises the possibility that genetic structure existed between Neandertals from different geographic regions and/or time points (1). Unaccounted-for population structure within the sample of Neandertals could inflate the estimate of effective population size (24). We therefore used PEC to retrieve a partial mtDNA genome (13,568 bp, SOM) from a second, more recent individual from Mezmaiskaya Cave [Mezmaiskaya 2, ~41,000 yr B.P. (15); table S2]. This mtDNA sequence unambiguously groups with the five complete mtDNAs from the western sites (fig. S13). Thus, no obvious phylogeographic structure can be discerned among the seven mtDNAs analyzed. Analysis of additional individuals from different times and locations will be necessary to examine broader patterns of temporal and/or geographic variation in Neandertal genetic diversity (1).

Low mtDNA diversity may reflect a low effective population size of Neandertals over a large part of their ~400,000-year history. Alternatively, it might be typical only of late Neandertals, who may have experienced a reduction in population size due to direct or indirect influences from modern humans expanding out of Africa. Of relevance for this question is the observation that the first complete Neandertal mtDNA genome had a higher rate of protein evolution (measured as the ratio of per site amino acid replacement to per site silent substitutions, $dN/dS$) than humans and great apes (5). This pattern is consistent with a long-term reduction in the effective population size of Neandertals, because purifying selection is expected to be less effective at purging weakly deleterious mutations in small populations (25). We tested this by partitioning $dN/dS$ into polymorphic and fixed differences along both the Neandertal and modern human lineages. $dN/dS$ for fixed differences was about twice as high along the Neandertal lineage (0.168) as along the modern human lineage (0.082; Fig. 2) or the lineages of chimpanzees and bonobos ($dN/dS < 0.08$; table S8). Furthermore, a likelihood model (26) in which $dN/dS$ is higher in Neandertals versus the other lineages best explains patterns of protein evolution across all groups (fig. S14). Although this difference appears pronounced, our estimate is based on relatively few amino acid substitutions overall, and lineage-specific variation in $dN/dS$ provides only an indirect assessment of long-term effective population size. Nonetheless, patterns of mtDNA protein evolution are consistent with the notion that the effective population size of Neandertals was small not only among late Neandertals but over a longer period of their existence. Clearly, the recovery of nuclear sequences from multiple Neandertals will be necessary to arrive at a more complete picture of their population history. Encouragingly, PEC and other targeted resequencing approaches (27–29) now make this feasible.



**Fig. 2.** mtDNA phylogeny of Neandertals and their close relatives. The top left panel gives the localities and estimated ages of the analyzed Neandertals. The phylogeny was estimated with a Bayesian approach (30) under a GTR+I+Γ model of sequence evolution. Modern humans are represented by the revised Cambridge reference sequence and a collection of 53 sequences sampled from major language groups. Nodes marked with an asterisk (*) have a posterior probability of 1.0. For clarity, support for major partitions within modern humans is not shown. Estimates of $dN/dS$ for 12 concatenated protein-coding genes are indicated below the branch.

**Table 1.** Mitochondrial DNA variation in Neandertals and modern humans.

| Sample | Length* | N† | Hap‡ | S§ | MPWD‖ | θ$_\pi$ (%)¶ |
|---|---|---|---|---|---|---|
| *All sites* | | | | | | |
| Neandertals | 16,565 | 6 | 5 | 55 | 20.4 | 0.123 |
| Modern humans | 16,547 | 54 | 54 | 658 | 60.5 | 0.365 |
| African | 16,556 | 21 | 21 | 365 | 76.5 | 0.462 |
| Non-African | 16,554 | 33 | 33 | 364 | 38.1 | 0.230 |
| European | 16,566 | 9 | 9 | 87 | 23.7 | 0.143 |
| European (expanded)# | 16,565 | 30 | 29 | 260 | 32.3 | 0.195 |
| *Third codon positions*** | | | | | | |
| Neandertals | 3,575 | 6 | 5 | 22 | 8.3 | 0.231 |
| Modern humans | 3,575 | 54 | 52 | 283 | 23.8 | 0.667 |
| African | 3,575 | 21 | 19 | 164 | 32.6 | 0.911 |
| Non-African | 3,575 | 33 | 33 | 134 | 13.5 | 0.378 |
| European | 3,575 | 9 | 9 | 23 | 5.9 | 0.165 |
| European (expanded)# | 3,575 | 30 | 29 | 91 | 10.5 | 0.293 |

*Number of aligned positions excluding alignment gaps.   †Number of sequences.   ‡Number of distinct mtDNA sequences (haplotypes).   §Number of variable sites.   ‖Mean number of pairwise differences.   ¶Average percentage of pairwise differences per site.   #Expanded to include additional, previously published complete mtDNA genomes. See SOM for details.   **Based on a 10,725-bp concatenated alignment of the 12 mtDNA genes coded on the same strand, excluding regions with overlapping reading frames.

**References and Notes**
1. L. Excoffier, *Curr. Biol.* **16**, R650 (2006).
2. J. Krause *et al.*, *Nature* **449**, 902 (2007).
3. D. Serre *et al.*, *PLoS Biol.* **2**, E57 (2004).
4. S. Paabo *et al.*, *Annu. Rev. Genet.* **38**, 645 (2004).
5. R. E. Green *et al.*, *Cell* **134**, 416 (2008).
6. M. T. Gilbert *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 8327 (2008).
7. M. T. Gilbert *et al.*, *Science* **320**, 1787 (2008).
8. C. Anderung, P. Persson, A. Bouwman, R. Elburg, A. Gotherstrom, *Forensic Sci. Int. Genet.* **2**, 104 (2008).
9. M. J. Blow *et al.*, *Genome Res.* **18**, 1347 (2008).
10. A. W. Briggs *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14616 (2007).
11. P. Brotherton *et al.*, *Nucleic Acids Res.* **35**, 5717 (2007).
12. R. Mackelprang, E. M. Rubin, *Science* **321**, 211 (2008).
13. R. W. Schmitz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13342 (2002).
14. C. Lalueza-Fox *et al.*, *Mol. Biol. Evol.* **22**, 1077 (2005).
15. A. R. Skinner *et al.*, *Appl. Radiat. Isot.* **62**, 219 (2005).
16. M. Hofreiter, V. Jaenicke, D. Serre, A. Haeseler Av, S. Paabo, *Nucleic Acids Res.* **29**, 4793 (2001).
17. I. V. Ovchinnikov *et al.*, *Nature* **404**, 490 (2000).
18. R. M. Andrews *et al.*, *Nat. Genet.* **23**, 147 (1999).
19. M. Ingman, H. Kaessmann, S. Pääbo, U. Gyllensten, *Nature* **408**, 708 (2000).
20. C. N. K. Anderson, U. Ramakrishnan, Y. L. Chan, E. A. Hadly, *Bioinformatics* **21**, 1733 (2005).
21. L. Excoffier, J. Novembre, S. Schneider, *J. Hered.* **91**, 506 (2000).
22. A. J. Drummond, A. Rambaut, *BMC Evol. Biol.* **7**, 214 (2007).
23. Q. D. Atkinson, R. D. Gray, A. J. Drummond, *Mol. Biol. Evol.* **25**, 468 (2008).

24. F. F. Jesus, J. F. Wilkins, V. N. Solferini, J. Wakeley, *Genet. Mol. Res.* **5**, 466 (2006).
25. T. Ohta, *Nature* **246**, 96 (1973).
26. S. L. Kosakovsky Pond, S. D. W. Frost, *Mol. Biol. Evol.* **22**, 478 (2004).
27. A. Gnirke et al., *Nat. Biotechnol.* **27**, 182 (2009).
28. E. Hodges et al., *Nat. Genet.* **39**, 1522 (2007).
29. J. P. Noonan et al., *Science* **314**, 1113 (2006).
30. F. Ronquist, J. P. Huelsenbeck, *Bioinformatics* **19**, 1572 (2003).
31. We thank P. Johnson, M. Knapp, A.-S. Malaspinas, M. Meyer, J. Sullivan, M. Slatkin, and anonymous reviewers

# The Human SepSecS-tRNA^Sec Complex Reveals the Mechanism of Selenocysteine Formation

Sotiria Palioura,[1] R. Lynn Sherrer,[1] Thomas A. Steitz,[1,2,3] Dieter Söll,[1,2]* Miljan Simonović[4]*

Selenocysteine is the only genetically encoded amino acid in humans whose biosynthesis occurs on its cognate transfer RNA (tRNA). *O*-Phosphoseryl-tRNA:selenocysteinyl-tRNA synthase (SepSecS) catalyzes the final step of selenocysteine formation by a poorly understood tRNA-dependent mechanism. The crystal structure of human tRNA^Sec in complex with SepSecS, phosphoserine, and thiophosphate, together with in vivo and in vitro enzyme assays, supports a pyridoxal phosphate–dependent mechanism of Sec-tRNA^Sec formation. Two tRNA^Sec molecules, with a fold distinct from other canonical tRNAs, bind to each SepSecS tetramer through their 13–base pair acceptor-TΨC arm (where Ψ indicates pseudouridine). The tRNA binding is likely to induce a conformational change in the enzyme's active site that allows a phosphoserine covalently attached to tRNA^Sec, but not free phosphoserine, to be oriented properly for the reaction to occur.

The 21st amino acid, selenocysteine (Sec), is distinct from other amino acids not only because it lacks its own tRNA synthetase, but also because it is the only one that is synthesized on the cognate tRNA in all domains of life [reviewed in (*1–4*)] in a process that is reminiscent of the tRNA-dependent synthesis of glutamine, asparagine, and cysteine in prokaryotes (*4*). The importance of Sec is illustrated by the embryonic lethal phenotype of the tRNA^Sec knockout mouse (*5*) and by the presence of Sec in the active sites of enzymes involved in removing reactive oxidative species and in thyroid hormone activation (*6, 7*). It is intriguing that the codon for

Sec is UGA, which is normally a translational stop signal (*1*). During translation of selenoprotein mRNAs, UGA is recoded by the interaction of a specialized elongation factor, SelB in bacteria and EFsec in humans, with a downstream Sec-insertion sequence element that forms a stem loop (*1, 3*).

The first step in Sec formation involves the misacylation of tRNA^Sec by seryl-tRNA synthetase (SerRS) to give Ser-tRNA^Sec [reviewed in (*3, 8*)]. Although the tertiary structure of tRNA^Sec was unknown, it was proposed that the mischarging reaction is possible because of similarities between the tRNA^Sec and tRNA^Ser structures (*3, 8*). In archaea and eukaryotes, the γ-hydroxyl group of Ser-tRNA^Sec is subsequently phosphorylated by *O*-phosphoseryl–tRNA kinase (PSTK) (*9*) to give *O*-phosphoseryl–tRNA^Sec (Sep-tRNA^Sec), which is then used as a substrate for the last synthetic

[1]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA. [2]Department of Chemistry, Yale University, New Haven, CT 06520, USA. [3]Howard Hughes Medical Institute, Yale University, New Haven, CT06520, USA. [4]Department of Biochemistry and Molecular Genetics, University of Illinois at Chicago, Chicago, IL 60607, USA.

*To whom correspondence should be addressed. E-mail: dieter.soll@yale.edu (D.S.); msimon5@uic.edu (M.S.)
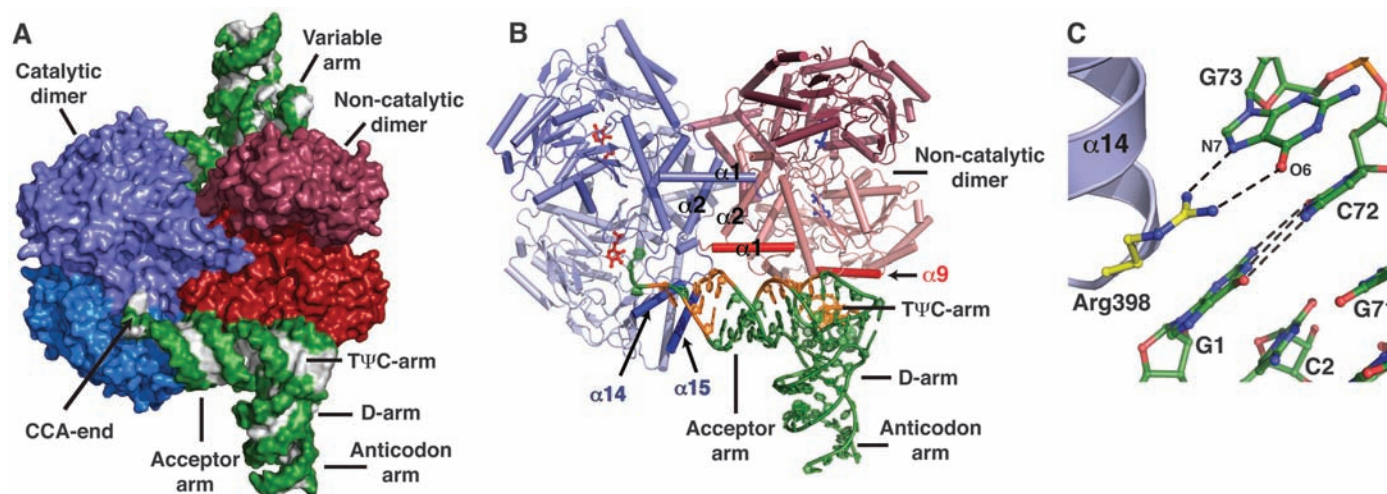
**Fig. 1.** Structure of human SepSecS in complex with unacylated tRNA^Sec. (**A**) Surface representation of the physiological complex of SepSecS with tRNA^Sec. The subunits of the catalytic dimer are dark and light blue, those of the noncatalytic dimer are dark and light red; the backbone and the bases of tRNA^Sec are green and gray, respectively. (**B**) The catalytic dimer interacts with the acceptor arm of tRNA^Sec through helices α14 and α15 (blue). The α1 helix (red) of the noncatalytic dimer interacts with the rest of the acceptor-TΨC arm. The regions of tRNA^Sec that interact with SepSecS are shown in orange; the rest is green. One tRNA^Sec molecule is shown for clarity. (**C**) Interactions between the discriminator base G73 and the conserved Arg[398] in the α14-β11 loop. The protein side chains are gold, and tRNA^Sec is green.

enzyme, SepSecS (8, 10). SepSecS catalyzes the conversion of the phosphoseryl moiety into the selenocysteinyl group by using selenophosphate as the selenium donor. An early observation that autoantibodies isolated from patients with type I autoimmune hepatitis targeted a ribonucleoprotein complex containing tRNA[Sec] led to the identification and characterization of the archaeal and the human SepSecS (2, 11). The crystal structures of the archaeal and murine SepSecS apo-enzymes and phylogenetic analysis suggested that SepSecS forms its own branch in the family of fold-type I pyridoxal phosphate (PLP) enzymes that goes back to the last universal common ancestor (12, 13).
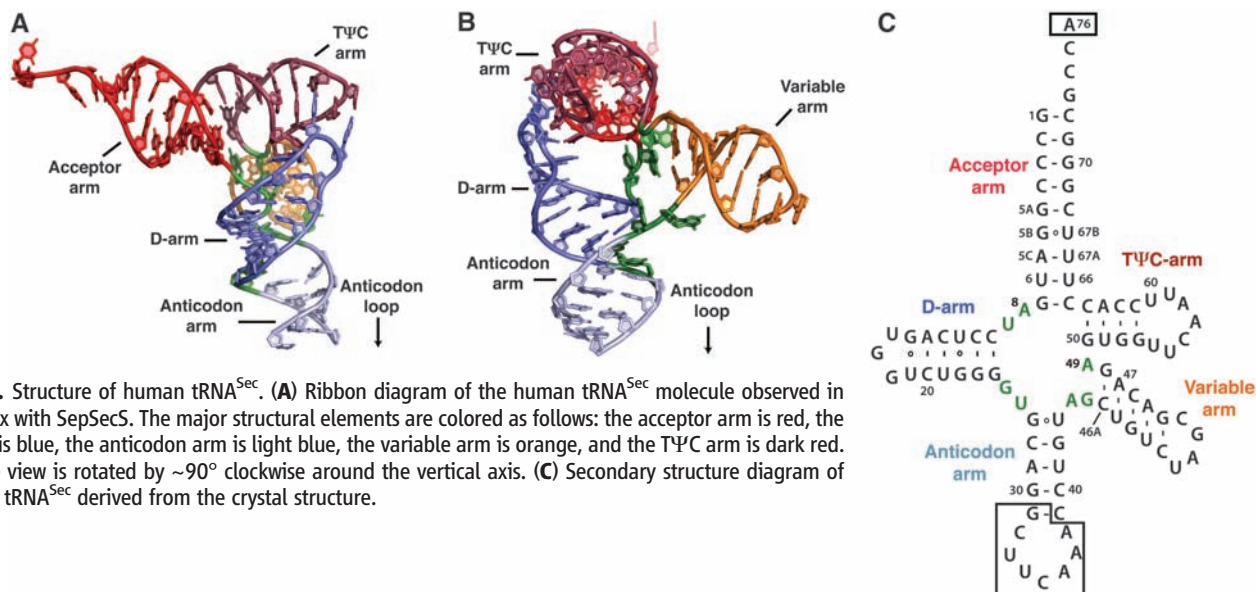


**Fig. 2.** Structure of human tRNA[Sec]. (**A**) Ribbon diagram of the human tRNA[Sec] molecule observed in complex with SepSecS. The major structural elements are colored as follows: the acceptor arm is red, the D-arm is blue, the anticodon arm is light blue, the variable arm is orange, and the TΨC arm is dark red. (**B**) The view is rotated by ~90° clockwise around the vertical axis. (**C**) Secondary structure diagram of human tRNA[Sec] derived from the crystal structure.



**Fig. 3.** Ligand binding to the active sites in the SepSecS-tRNA[Sec] complex. In (A), (C), (D), (E), and (F), the catalytic PLP-monomer is purple, the P-loop monomer is light blue, Sep is gold, tRNA is green, and PLP is magenta. In (A) and (B), the unbiased omit electron density map (green mesh) is contoured at 3.5 σ. (**A**) Phosphoserine binds to the catalytic sites in two orientations (Sep[P] and Sep[N]). (**B**) Thiop (orange) binds only to the non-catalytic site. The PLP-monomer is brown, and the P-loop monomer is pink. (**C**) The Sep[N] amino group is ~12 Å away from the Schiff base. Arg[97], Gln[105], and Arg[313] coordinate phosphate, and the carboxyl group interacts with Lys[173]. (**D**) The amino group of Sep[P] interacts with Gln[172] and is 3.5 Å away from the Schiff base. The carboxyl group interacts with Gln[172], whereas Ser[98], Gln[105], and Arg[313] coordinate phosphate. (**E**) The P-loop adopts different conformation after tRNA[Sec] binding. The noncatalytic dimer is pink, the catalytic dimer is light blue. Steric clashes between the noncatalytic P-loop and Sep[N] are shown (double arrow). (**F**) A model of CCA-Sep[P] (gold) in the catalytic active site. A76 binds between the side chains of Arg[97] and Lys[173], whereas C75 interacts with Arg[97].
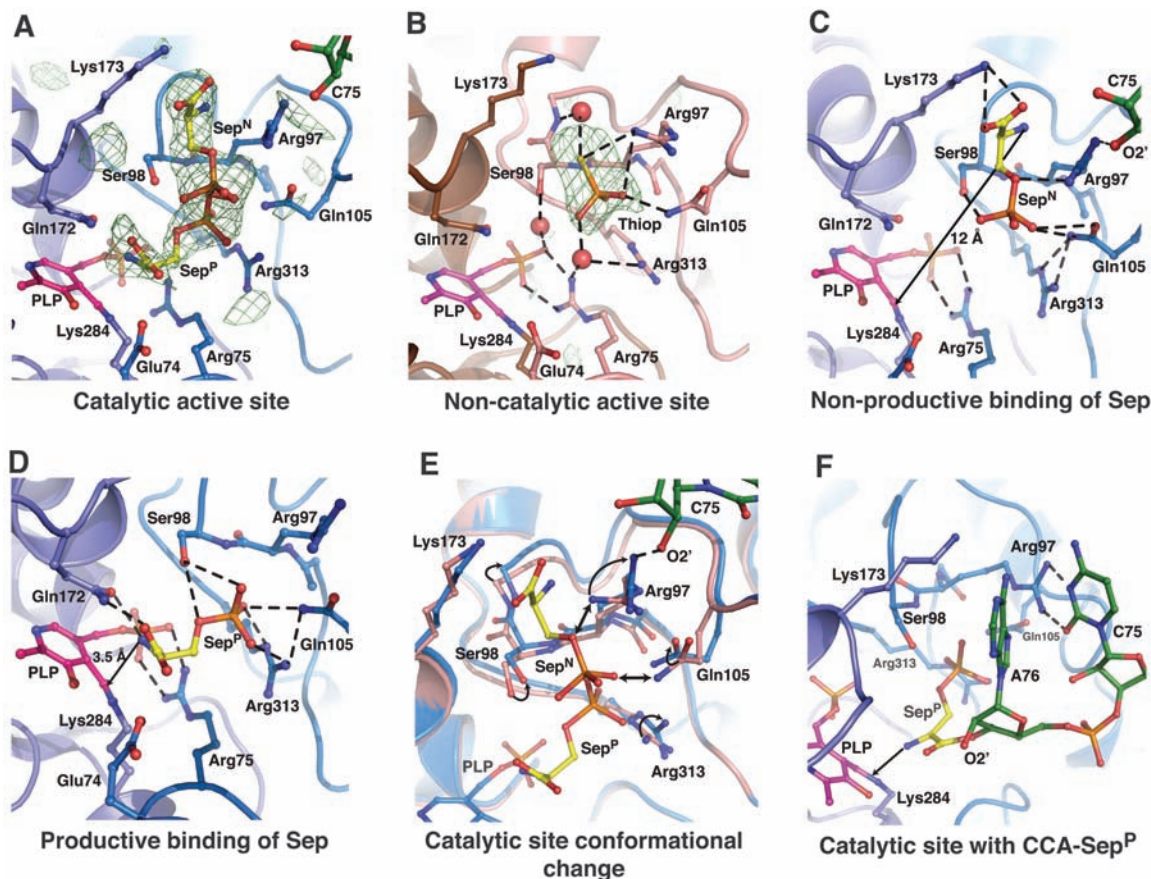
In contrast to its closest homologue, SepCysS, which functions as a dimer to convert Sep-tRNA$^{Cys}$ to Cys-tRNA$^{Cys}$ in archaea (14, 15), SepSecS forms a stable tetramer (12, 13). SepSecS acts on phosphoserine that is linked to tRNA$^{Sec}$ and not on free phosphoserine or Ser-tRNA$^{Sec}$ (8, 13). However, the molecular basis for substrate discrimination and the roles of PLP and tRNA$^{Sec}$ in the mechanism of Sep to Sec conversion are not clear. To explore these questions, we have determined the crystal structure of the quaternary complex between human SepSecS, unacylated tRNA$^{Sec}$, and a mixture of O-phosphoserine (Sep) and thio-

phosphate (Thiop) to 2.8 Å resolution. The observed intensity divided by its standard deviation [$I/\sigma(I)$] of the x-ray diffraction data are 2 at 3.0 Å resolution, but the data out to 2.8 Å resolution, where $I/\sigma(I)$ drops to 1, were included in the structure refinement.

Human SepSecS forms a tetramer that is bound to two tRNA$^{Sec}$ molecules in the crystal (Fig. 1, A and B, and fig. S1). Computational modeling suggests that the tetrameric enzyme could potentially bind up to four tRNA$^{Sec}$ molecules (fig. S2), but these additional tRNA-binding sites are blocked by crystal packing in

our crystal (16). Electron density for the 23 N-terminal and 15 C-terminal residues of SepSecS, as well as for the anticodon loop (nucleotides 31 to 38) and A76 of tRNA$^{Sec}$, was of poor quality, and these residues were not included in the final model. Each SepSecS monomer has a PLP cofactor covalently linked to the Nε-amino group of the conserved Lys$^{284}$ by means of formation of a Schiff base (internal aldimine). Two SepSecS monomers form a homodimer, and two active sites are formed at the dimer interface. The two homodimers associate into a tetramer through interactions between the N-terminal α1-loop-α2
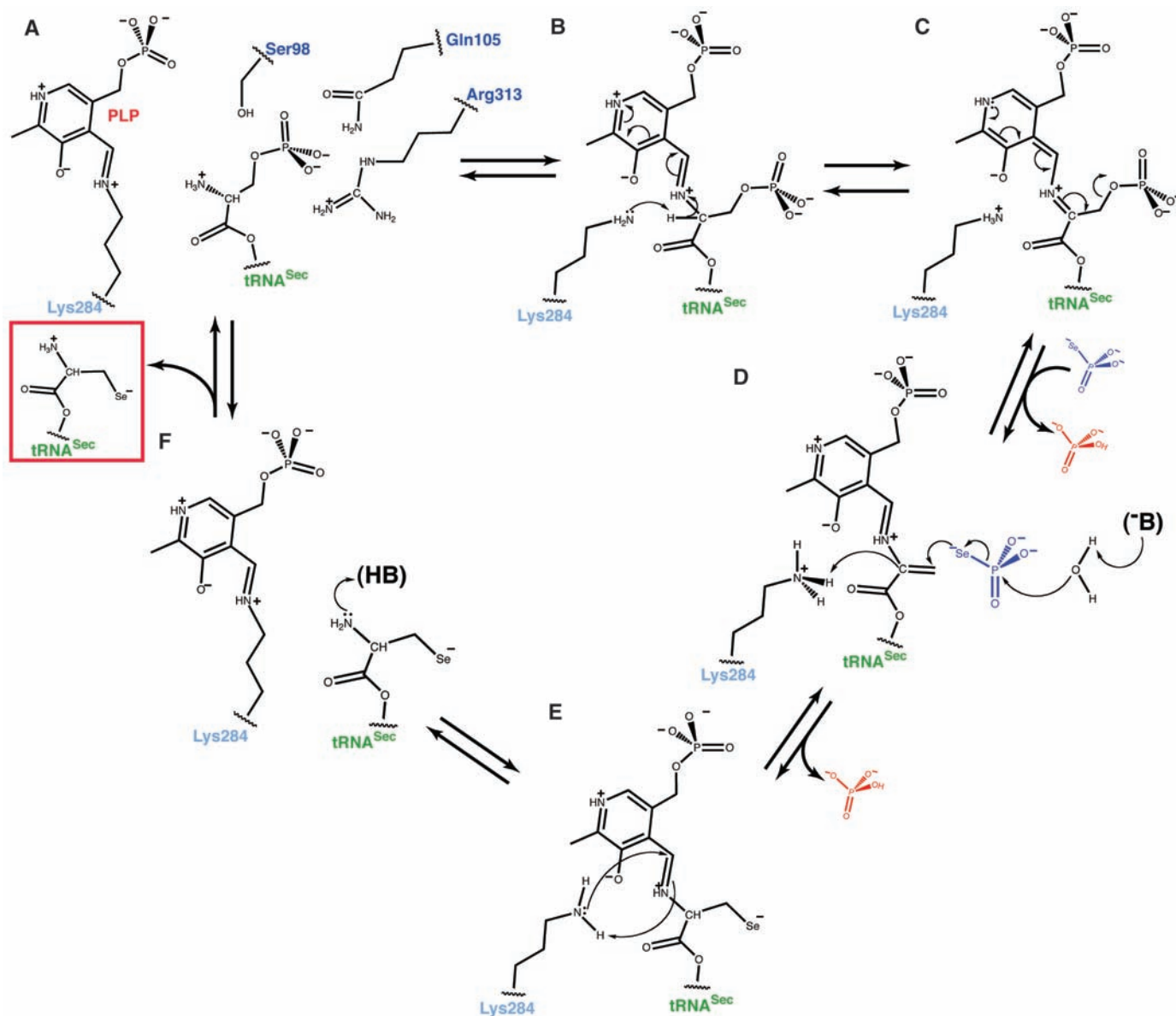


**Fig. 4.** The PLP-dependent mechanism of Sep to Sec conversion. (**A**) The phosphoseryl moiety of Sep-tRNA$^{Sec}$ is bound to the active site similar to Sep$^P$. The amino group is oriented for attack on the Schiff base, whereas the phosphoryl group is stabilized by the side chains of Ser$^{98}$, Gln$^{105}$, and Arg$^{313}$. Hydrogen bonds are shown in dashed lines. (**B**) After the formation of the external aldimine, the side chain of Lys$^{284}$ rearranges and abstracts the Cα proton from Sep. The protonated pyridine ring of PLP stabilizes the carbanion. (**C**) Electron delocalization leads to a rapid β-elimination of phosphate and to the formation

of dehydroalanyl-tRNA$^{Sec}$. Free phosphate dissociates, and selenophosphate binds to the active site. (**D**) An unidentified base ($^-$B) activates water that hydrolyzes selenophosphate. Free phosphate dissociates again, and selenium attacks the dehydroalanyl-tRNA$^{Sec}$. Lys$^{284}$ returns the proton to the Cα carbon, and the selenocysteinyl moiety is formed. (**E**) The reaction of reverse transaldimination is shown. Lys$^{284}$ forms the Schiff base, with PLP leading to a release of the oxidized form of Sec-tRNA$^{Sec}$ (red box). (**F**) The free amino group of Sec-tRNA$^{Sec}$ is protonated, and the active site of SepSecS is regenerated.

motifs (Fig. 1B). Given that SepCysS is a dimer and that the active sites of one homodimer do not communicate with the active sites of the other homodimer in the apo-SepSecS tetramer (12, 13), the reason for the tetrameric organization of SepSecS was not known. The mode of tRNA$^{Sec}$ binding to SepSecS provides an answer to this question.

The CCA ends of both tRNA$^{Sec}$ molecules point to the active sites of the same homodimer, which we shall refer to as the catalytic dimer (Fig. 1A). The other homodimer, which we shall refer to as the noncatalytic dimer, serves as a binding platform that orients tRNA$^{Sec}$ for catalysis (Fig. 1A). The structure reveals that SepSecS binds only to the acceptor-, TΨC-, and variable arms of tRNA$^{Sec}$ (Fig. 1B and fig. S3A). (Ψ, pseudouridine.) The tip of the acceptor arm interacts with the C terminus of the catalytic dimer, whereas the rest of the acceptor-, TΨC-, and variable arms wrap around a monomer from the noncatalytic dimer (Fig. 1, A and B, and fig. S3A). The most important binding element is an interaction between the discriminator base G73 of tRNA$^{Sec}$ and the conserved Arg$^{398}$ of the catalytic dimer (Fig. 1C). The guanidinium group of Arg$^{398}$ forms hydrogen bonds with the Hoogsteen face of G73. The discriminator base G73 of tRNA$^{Sec}$ is universally conserved in archaea and eukaryotes. Neither adenine nor cytidine in position 73 could form hydrogen bonds with Arg$^{398}$ because they have amino groups instead of the keto group. Also, if a cytidine or uridine were in position 73, the C5 and C6 atoms of the pyrimidine ring would clash with the side chain of Thr$^{397}$ and thus prevent the interaction of these bases with Arg$^{398}$.

Moreover, the side chain of Lys$^{463}$ from the C-terminal helix α15 forms a hydrogen bond with the backbone oxygen of G69 from the acceptor arm. We propose that autoantibodies bind to an interface that lies between the α15 helix of SepSecS and the tip of the acceptor arm of tRNA$^{Sec}$ (Fig. 1B), which enables them to precipitate the entire ribonucleoprotein complex (11). The interaction of the antibody with a region that lies close to the acceptor stem–active site interface may inhibit the function of SepSecS. This would be similar to the mechanism of autoimmune hepatitis type 2, where LKM-1 autoantibodies (anti–liver-kidney microsomal antibodies) inhibit cytochrome P450 isoenzyme 2D6 (CYP2D6) and contribute to the pathogenesis of the disease (17).

The tRNA$^{Sec}$ molecule is anchored to the SepSecS tetramer, on one end by interactions between the tip of its acceptor arm (G73) and the catalytic dimer (Arg$^{398}$) (Fig. 1C) and on the other end by interactions between its variable arm (C46L) and the noncatalytic dimer (Arg$^{271}$) (fig. S3B). The rest of the interactions between the α1 helix of the noncatalytic dimer of SepSecS (Arg$^{26}$, Lys$^{38}$, and Lys$^{40}$) and the acceptor-TΨC arm (C2, G50, and C64) further stabilize the complex (fig. S3C). For instance, the side chain of Arg$^{26}$ forms hydrogen bonds with the phosphate backbone of C2-C3 of the acceptor arm

and the Nε-amino group of Lys$^{38}$ is within the hydrogen bonding distance from both the O2′-hydroxyl group and the O2 atom of C64 (fig. S3C), whereas the side chain of Lys$^{40}$ interacts with the phosphate oxygen of G50 from the TΨC arm (fig. S3C). The interactions between SepSecS and tRNA$^{Sec}$ seen in the crystal are consistent with in vivo activity assays of SepSecS mutants (fig. S4A) (16). For example, replacing Arg$^{398}$ with either alanine or glutamate renders the enzyme completely inactive, which suggests that the interaction between the discriminator base and the highly conserved Arg$^{398}$ of the catalytic dimer is critical for tRNA$^{Sec}$ recognition (fig. S4A).

Human tRNA$^{Sec}$ contains 90 nucleotides rather than the conventional 75 nucleotides of canonical tRNA molecules. The structure shows that human tRNA$^{Sec}$ adopts a unique 9/4 fold with a 13–base pair (bp) acceptor-TΨC arm (where 9 and 4 reflect the number of base pairs in the acceptor and TΨC arms, respectively) and a long variable arm (Fig. 2) (16). This resolves a controversy between conflicting models in which the 7/5 model suggested 12 bp in the acceptor-TΨC arm, as found in all known tRNA structures (18), whereas the 9/4 model suggested a unique 13-bp acceptor-TΨC arm (19). The 13-bp acceptor-TΨC arm and a long variable arm are distinct structural features that serve as major recognition motifs for binding to SepSecS (Fig. 1, A and B, and fig. S3). Indeed, based on our modeling analysis with tRNA$^{Asp}$, we suggest that tRNA$^{Ser}$, which contains both the G73 discriminator base and a long variable arm, is not able to bind to SepSecS because of its shorter acceptor-TΨC arm (fig. S5, A, B, and C). This is reminiscent of the rejection of tRNA$^{Sec}$ by bacterial EF-Tu (20). However, modeling of the SerRS-tRNA$^{Sec}$ complex based on the crystal structure of the bacterial SerRS in complex with tRNA$^{Ser}$ (21) suggests that SerRS can recognize the variable arm of tRNA$^{Sec}$ similarly to tRNA$^{Ser}$, which provides an explanation for the inability of SerRS to discriminate between tRNA$^{Ser}$ and tRNA$^{Sec}$ (fig. S5D). Finally, the observations that the archaeal PSTK and SepSecS can act on E. coli tRNA$^{Sec}$ in vivo; that human SepSecS can use E. coli tRNA$^{Sec}$ in vivo, as well as the archaeal tRNA$^{Sec}$ in vitro (8); and that the bacterial selenocysteine synthase can use both archaeal and murine tRNA$^{Sec}$ in vitro (10) suggest that the length of the acceptor-TΨC arm of tRNA$^{Sec}$, the position of the variable arm, and the mode of tRNA$^{Sec}$ recognition are likely to be conserved in all domains of life.

To explore the mechanism by which the phosphoryl group is converted to the selenocysteinyl moiety, we soaked crystals of the binary SepSecS-tRNA$^{Sec}$ complex in a solution containing a mixture of O-phosphoserine and thiophosphate. Sep and Thiop were used as mimics of the phosphoseryl group attached to tRNA$^{Sec}$ and selenophosphate, respectively. Although both ligands were used in the soaking experiments, Sep bound only to the active sites of the

catalytic dimer (Fig. 3A), whereas Thiop bound only to the active sites of the noncatalytic dimer (Fig. 3B), i.e., Sep can bind to an active site only in the presence of tRNA$^{Sec}$. Both Thiop and the phosphoryl group of Sep bind to the same binding pocket (Fig. 3, A and B), which suggests that a specific active site can accommodate only one ligand at a time. Sep binds to the catalytic active site in either of two different orientations (Fig. 3A). The phosphoryl group occupies a similar location in the two orientations, whereas the seryl moieties are rotated by ~90° around the phosphate group. In the nonproductive orientation of Sep (Sep$^N$), its seryl group is sandwiched between the side chains of Arg$^{97}$ and Lys$^{173}$ at the catalytic dimer interface, whereas its amino group lies ~12 Å away from the PLP Schiff base (Fig. 3C). Thus, it is unlikely that SepSecS would act on Sep$^N$, unless a PLP-independent mechanism is utilized. In the productive-like orientation of Sep (Sep$^P$), the amino group of Sep is ~3.5 Å away from the Schiff base, yet it is not positioned for attack because of a hydrogen bond with Gln$^{172}$ (Fig. 3D). The carboxyl group of Sep$^P$ also forms a hydrogen bond with the side chain of Gln$^{172}$, whereas its phosphoryl group anchors the ligand into the active site through its interactions with the side chains of Ser$^{98}$, Gln$^{105}$, and Arg$^{313}$. That the phosphoryl group is required to properly position the ligand for catalysis explains why the obligate substrate for SepSecS is Sep-tRNA$^{Sec}$ and not Ser-tRNA$^{Sec}$, which is the physiological substrate for the bacterial selenocysteine synthase.

A comparison of the catalytic and noncatalytic sites reveals that Sep binds to the active site only in the presence of tRNA$^{Sec}$ because the conformation of the P-loop (residues Gly$^{96}$ to Lys$^{107}$) differs between the two active sites (Fig. 3E). In the noncatalytic dimer, the guanidinium group of Arg$^{97}$ and the side chain of Gln$^{105}$ are rotated toward the phosphate-binding groove, where they coordinate Thiop (Fig. 3, B and E). In the catalytic dimer, the side chain of Arg$^{97}$ rotates away from the phosphate-binding groove and forms a hydrogen bond with the 2′-OH group of C75. Gln$^{105}$ also rotates away from the phosphate, and this concerted movement of Arg$^{97}$ and Gln$^{105}$ in the P-loop on tRNA$^{Sec}$ binding allows Sep to bind to the active site (Fig. 3E). Free Sep enters the active site through the gate formed by the side chains of Arg$^{97}$, Gln$^{105}$, and Lys$^{173}$ and gets trapped in the Sep$^N$ orientation. Electron density for both the amino and the carboxyl groups of Sep$^P$ is weak, which suggests higher mobility in this part of Sep$^P$, presumably because of free rotation around the Cα-Cβ bond. This explains why the free amino group of Sep$^P$ cannot be positioned appropriately for attack onto the Schiff base and why the reaction does not occur in the crystal. Thus, the covalent attachment of Sep to tRNA$^{Sec}$ is necessary for the proper placement of the Sep moiety into the active site and for orienting the amino group of Sep for attack onto the Schiff base of PLP (Fig. 3F).

We used both in vivo and in vitro activity assays to investigate the mechanism of Sep-tRNA to Sec-tRNA conversion by human SepSecS. First, the reduction of the Schiff base by sodium borohydride to form a chemically stable secondary amine and thus to cross-link PLP to Lys[284] renders SepSecS completely inactive in vitro (fig. S4B). The catalytic activity of SepSecS is also quenched on removal of PLP by treatment with hydroxylamine (fig. S4B). Some residual activity that is observed after hydroxylamine treatment is probably because of incomplete removal of PLP (fig. S4B). Second, we show that the Arg75Ala, Gln105Ala, and Arg313Ala mutants are inactive in vivo (fig. S4A). These residues are involved in coordinating either the phosphate group of PLP or that of Sep. Finally, the in vivo activities of the Arg97Ala, Arg97Gln, Lys173Ala, and Lys173Met mutants are indistinguishable from that of the wild-type enzyme, which confirms that Arg[97] and Lys[173] are involved only in the nonproductive binding of free Sep (fig. S4A).

On the basis of our findings, we propose the following PLP-based mechanism of Sep-tRNA to Sec-tRNA conversion. The reaction begins by the covalently attached Sep being brought into the proximity of the Schiff base when Sep-tRNA[Sec] binds to SepSecS. The amino group of Sep can then attack the Schiff base formed between Lys[284] and PLP, which yields an external aldimine (Fig. 4, A and B). The reoriented side chain of Lys[284] abstracts the Cα proton from Sep (Fig. 4C), and the electron delocalization by the pyridine ring assists in rapid β-elimination of the phosphate group, which produces an intermediate dehydroalanyl-tRNA[Sec] (Fig. 4, C and D). After phosphate dissociation and binding of selenophosphate, the concomitant attack of water on the selenophosphate group and of the nucleophilic selenium onto

the highly reactive dehydroalanyl moiety yield an oxidized form of Sec-tRNA[Sec] (Fig. 4D). The protonated Lys[284], returns the proton to the Cα carbon and then attacks PLP to form an internal aldimine (Fig. 4E). Finally, Sec-tRNA[Sec] is released from the active site (Fig. 4F).

This mechanism is clearly distinct from the persulfide-intermediate mechanism in the Sep-tRNA[Cys] to Cys-tRNA[Cys] reaction (22) and explains why SepSecS does not group together with its closest homolog, SepCysS, in the family tree of fold-type I PLP enzymes (12). Moreover, the proposed mechanism for SepSecS is similar to the one used by the bacterial SelA that also proceeds through a dehydroalanyl-tRNA[Sec] intermediate (23). SepSecS therefore uses a primordial tRNA-dependent catalytic mechanism in which the PLP cofactor is directly involved, while using a tetrameric fold-type I architecture as the scaffold for binding the distinct structure of tRNA[Sec].

### References and Notes

1. A. Böck, M. Thanbichler, M. Rother, A. Resch, in *Aminoacyl-tRNA Synthetases,* M. Ibba, C. Francklyn, S. Cusack, Eds. (Landes Bioscience, Georgetown, TX, 2005), pp. 320–327.
2. D. Su *et al.*, *IUBMB Life* **61**, 35 (2009).
3. A. Ambrogelly, S. Palioura, D. Söll, *Nat. Chem. Biol.* **3**, 29 (2007).
4. K. Sheppard *et al.*, *Nucleic Acids Res.* **36**, 1813 (2008).
5. M. R. Bösl, K. Takaku, M. Oshima, S. Nishimura, M. M. Taketo, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5531 (1997).
6. M. P. Rayman, *Lancet* **356**, 233 (2000).
7. G. V. Kryukov *et al.*, *Science* **300**, 1439 (2003).
8. J. Yuan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18923 (2006).
9. B. A. Carlson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12848 (2004).
10. X. M. Xu *et al.*, *PLoS Biol.* **5**, e4 (2007).
11. C. Gelpi, E. J. Sontheimer, J. L. Rodriguez-Sanchez, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9739 (1992).
12. Y. Araiso *et al.*, *Nucleic Acids Res.* **36**, 1187 (2008).
13. O. M. Ganichkin *et al.*, *J. Biol. Chem.* **283**, 5849 (2008).
14. A. Sauerwald *et al.*, *Science* **307**, 1969 (2005).
15. R. Fukunaga, S. Yokoyama, *J. Mol. Biol.* **370**, 128 (2007).
16. Materials and methods are available as supporting material on *Science* Online.
17. J. Herkel, M. P. Manns, A. W. Lohse, *Hepatology* **46**, 275 (2007).
18. A. Ioudovitch, S. V. Steinberg, *RNA* **4**, 365 (1998).
19. C. Sturchler, E. Westhof, P. Carbon, A. Krol, *Nucleic Acids Res.* **21**, 1073 (1993).
20. J. Rudinger, R. Hillenbrandt, M. Sprinzl, R. Giege, *EMBO J.* **15**, 650 (1996).
21. V. Biou, A. Yaremchuk, M. Tukalo, S. Cusack, *Science* **263**, 1404 (1994).
22. S. I. Hauenstein, J. J. Perona, *J. Biol. Chem.* **283**, 22007 (2008).
23. K. Forchhammer, A. Böck, *J. Biol. Chem.* **266**, 6324 (1991).
24. We thank C. Axel Innis, Gregor Blaha, Robin Evans, Michael Strickler, Michael E. Johnson, Bernard D. Santarsiero, Jimin Wang, and the NE-CAT beamline staff (APS, ANL, Chicago) for their help during data collection and structure determination. We thank Dan Su, Theodoros Rampias, and Kelly Sheppard for helpful discussions. Atomic coordinates and structure factors have been deposited in the Protein Data Bank (code 3HL2). Supported by grants from DOE (to D.S.) and NIGMS (to T.A.S. and D.S.). In the initial phase of this study M.S. was supported by HHMI at Yale University. S.P. holds a fellowship of the Yale University School of Medicine MD/PhD Program. R.L.S. was supported by a Ruth L. Kirschstein National Research Service Award. Authors' contributions: S.P. designed the research, collected and analyzed the data, and wrote the manuscript; R.L.S. did research and read the manuscript; T.A.S. designed the research and wrote the manuscript; D.S. designed the research and wrote the manuscript; M.S. designed the research, collected and analyzed the data, and wrote the manuscript.

# Tiger Moth Jams Bat Sonar

Aaron J. Corcoran,[1]* Jesse R. Barber,[2] William E. Conner[1]*

In response to sonar-guided attacking bats, some tiger moths make ultrasonic clicks of their own. The lepidopteran sounds have previously been shown to alert bats to some moths' toxic chemistry and also to startle bats unaccustomed to sonic prey. The moth sounds could also interfere with, or "jam," bat sonar, but evidence for such jamming has been inconclusive. Using ultrasonic recording and high-speed infrared videography of bat-moth interactions, we show that the palatable tiger moth *Bertholdia trigona* defends against attacking big brown bats (*Eptesicus fuscus*) using ultrasonic clicks that jam bat sonar. Sonar jamming extends the defensive repertoire available to prey in the long-standing evolutionary arms race between bats and insects.

The ability to pinpoint airborne insects in darkness, by echolocation (1), allowed bats to master nocturnal insectivory (2) and set the stage for the evolution of defensive countermeasures by insect prey (3). Some insects gained ears (4) and evasive maneuvering (5). Tiger moths (Lepidoptera: Arctiidae) developed the ability to click ultrasonically in response to attacking bats (6, 7). Decades of research on moth click de-

fenses have led to three, not mutually exclusive, hypotheses regarding their function—startle (8), acoustic aposematism ("warning") (8–14), and sonar interference ("jamming") (15–17). When ultrasonic clicks are paired with unpalatable prey, bats learn to perceive clicks as a warning of unprofitability (8, 10–14). Moth clicks also startle inexperienced bats or bats that have not heard clicks for multiple days (8). Because bats habit-

uate to startle quickly, its effectiveness as a defense requires clicking moths to be rare. This situation does not appear typical in nature (12).

Finally, moth clicks may disrupt the sonar of an attacking bat (15–17). Clicks might diminish a bat's acuity in determining target distance (17–19) or feign echoes from objects that do not exist (15). However, evidence that moth clicks can disrupt bat attacks by jamming sonar is lacking. One recent study found that moth clicks had no discernible effect on attacking bats unless clicks were paired with defensive chemistry (13). All previous studies, however, tested moths with relatively low duty cycles, or sound production per unit time (20). High-duty-cycle moth clicks, such as those of *Bertholdia trigona* (Fig. 1 and

[1]Department of Biology, Wake Forest University, Winston-Salem, NC 27106, USA. [2]Department of Fish, Wildlife and Conservation Biology, Colorado State University, Fort Collins, CO 80523, USA.

*To whom correspondence should be addressed. E-mail: corcaj8@wfu.edu (A.J.C.); conner@wfu.edu (W.E.C.)

table S2), are more likely to jam bats (*17, 18*), including the big brown bat (*Eptesicus fuscus*), which co-occurs with *B. trigona* in much of Central America and northward to Colorado, USA.

To determine whether the clicks of *B. trigona* startle, warn, or jam bats, we pitted *E. fuscus* against *B. trigona* in a flight room equipped with high-speed infrared video cameras and an ultrasonic microphone (*21*). The pattern of a naïve bat's success in catching clicking moths, when measured over several days, should differ for each of the three proposed moth click functions (Fig. 2A) (*13*). For a warning sound, bats initially capture and drop distasteful moths, then learn to abort attacks on hearing clicks (*8, 13, 22*). Alternatively, a bat could be initially startled by the clicks but would quickly habituate to the sounds (*8, 23*). A jamming defense, however, would immediately deter attack-

ing bats, and the effect of the defense would persist with time (Fig. 2A).

We trained three naïve juvenile *E. fuscus* (bats J1, J2, and J3) and one adult *E. fuscus* (bat A1) to capture tethered moths in a sound-attenuated flight room. Naïve bats allow us to control prior experience, whereas an adult bat's prior experience catching insects may allow it to better overcome moth defense. On each of nine consecutive nights, 16 tethered moths—4 *B. trigona*; 4 silent, palatable, size-matched noctuid novelty controls; and 8 silent, palatable wax moths (female *Galleria mellonella*, used in pre-experiment bat training)—were presented to each bat individually in a random order. Bats were allowed up to 1 min or five approaches for each moth. We presented unaltered *B. trigona* to the bats on nights one through seven and *B. trigona* lacking the ability to click (from tymbal abla-

tion) on nights eight and nine. Seventy-eight percent of the *B. trigona* (87 of 112) responded acoustically to bat sonar. The few intact *B. trigona* that did not produce sound in response to bat attack were excluded from the analysis. We used contact with the moths as our measure of attack success. Palatability of *B. trigona* to each bat was taken as the proportion of captured moths eaten rather than dropped. We conducted feeding trials with silenced moths and four additional stationary bats to further assess the palatability of *B. trigona*. Finally, we analyzed echolocation recordings to measure bat behavioral response to moth clicks (*21*).

In stationary feeding trials with silenced moths, three bats regularly ate *B. trigona* (J4, 92%, $n = 12$; J5, 92%, $n = 12$; J6, 83%, $n = 6$), and one did not (J7, 0%; $n = 8$). Similarly, of the bats capturing moths in flight, three regularly ate *B. trigona* (J1, 94%, $n = 16$; J2, 50%, $n = 14$; A1, 89%, $n = 18$), and one did not (J3, 0%, $n = 12$). The observed differences between individuals may reflect a greater reluctance of some individuals to incorporate new items into their diet, as has been reported for some birds (*24*). *B. trigona* appears palatable; the six bats ate the moths with no detectable ill effects. Because warning and jamming are not mutually exclusive defense functions (*25*), our ability to detect a jamming effect is diminished for the bat that rejected *B. trigona*—any moth avoidance could be caused by jamming, warning, or both. Thus, to unambiguously test the hypothesis of sonar jamming, we focus our analysis on the three bats that ate *B. trigona* in flight (J1, J2, and A1).

Each of the three bats contacted control moths over 400% more often than clicking *B. trigona* (Table 1 and movie S1). In contrast, bats contacted 100% of *B. trigona* when the moths' tymbals
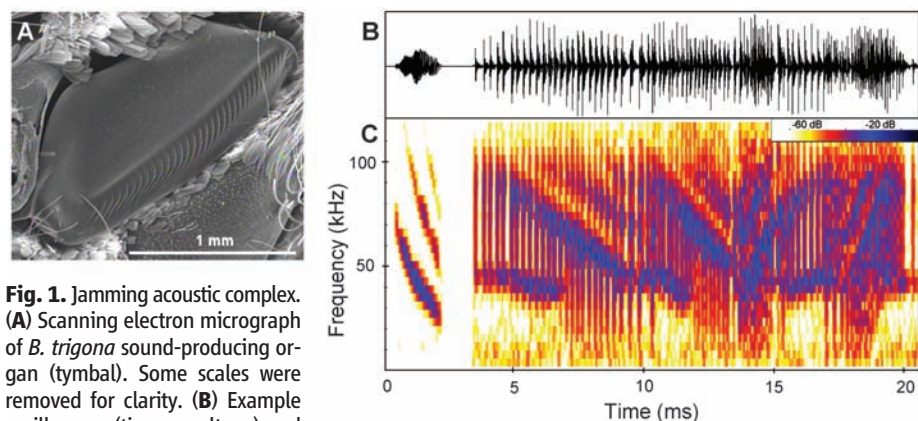
**Fig. 1.** Jamming acoustic complex. (**A**) Scanning electron micrograph of *B. trigona* sound-producing organ (tymbal). Some scales were removed for clarity. (**B**) Example oscillogram (time × voltage) and (**C**) spectrogram showing a simulated *E. fuscus* echo, reflected off of *B. trigona* at 1 m (left), and two overlapping *B. trigona* click modulation cycles (right). Echo and clicks are presented at average estimated intensities at the bat's location 1.0 m from the moth (*21*).

**Fig. 2.** Predicted and observed success rates for *E. fuscus* attacking moths using various defensive strategies. (**A**) Predicted bat attack success over successive nights for three proposed moth click functions: startle, aposematism, and jamming. (**B**) Observed bat attack success rates for three *E. fuscus* (J1, circles; J2, triangles; A1, squares) on *B. trigona* (closed symbols) and noctuid controls (open symbols). The solid line shows mean values for attacks on *B. trigona*. (**C**) Success rates of bats attacking intact (sound-producing) and silenced *B. trigona*. Means ± SEM. Some values were offset slightly for display.
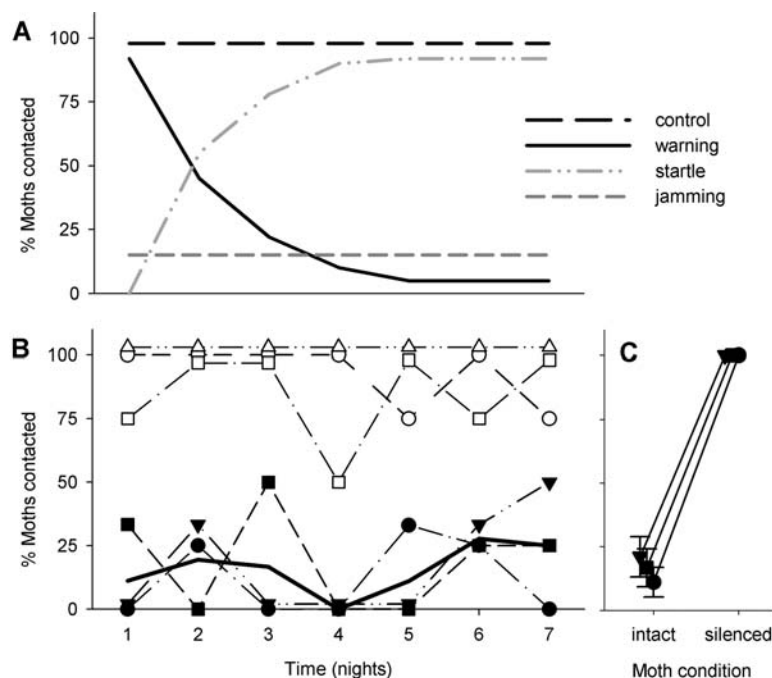
were ablated (Fig. 2C and movie S2). This confirms that the bats' poor performance on *B. trigona* was caused by the moth clicks. Rates of contact with sound-producing moths did not change during the experiment (Fig. 2B) [Friedman nonparametric analysis of variance (ANOVA), $\chi^2 = 2.63$, df = 6, $P = 0.85$]. Thus, prior learning was not required for the defense to be effective, nor did the bats habituate to the clicks. Bat contact rates did not improve on succeeding attacks on the same moth—attacks that often occurred within seconds of each other (table S1) (Friedman nonparametric ANOVA, $\chi^2 = 2.55$, df = 4, $P = 0.64$). This renders startle unlikely because nearly all mammals, including bats, habituate to startle (*8*, *23*). Echolocation recordings (Fig. 3) showed bats persisting in attack attempts after onset of moth clicks to varying degrees throughout the experiment (fig. S1A). In contrast, bats in a previous study quickly learned to abort attacks on the noxious *Cycnia tenera* (fig. S1B) (*22*). This further suggests that the bats in our study did not learn to take the moth clicks as a warning—the bats continued attacks despite hearing the sounds. It also indicates that moth clicks are not generally aversive to bats [supporting online material (SOM) text]. Together, we

take the above results as strong evidence for a jamming function of the moth sound. The defense was effective immediately and persistently on juvenile and adult *E. fuscus* (Fig. 2B) that frequently tried to capture the clicking moths (fig. S1A) but had much difficulty doing so.

We also observed atypical echolocation behavior in response to *B. trigona* clicks. Each bat attacked all control moths by progressing through the approach, track, and terminal phases [e.g., (*26*, *27*)]. However, in about a third (J1, 21%, $n = 19$; J2, 40%, $n = 15$; A1, 28%, $n = 25$) of attacks on *B. trigona*, the bat reversed the attack phase from track to approach or terminal to track, and then continued the attack (Fig. 3). This behavior did not diminish with experience (Friedman nonparametric ANOVA, $\chi^2 = 4.5$, df = 6, $P = 0.61$), as would be expected with startle. It appears to be a direct response of the bats to the sonar interference caused by the moth clicks.

The ultrasonic clicks of *B. trigona* appear well-suited for jamming echolocation (SOM text). Under all proposed jamming mechanisms, high-duty-cycle sounds made by highly developed tymbals (Fig. 1) should better jam echolocation (*15*, *17–19*). Empirical evidence now supports

this prediction; the high-duty-cycle sounds of *B. trigona* jammed bats, whereas the low-duty-cycle sounds of *Euchaetes egle* were ineffective jamming signals under nearly identical conditions (*13*). The use of tiger moth sound for aposematic signaling requires only simple tymbal structure and appears widespread in the Arctiidae (*7*, *10–13*). Thus, acoustic aposematism may have been a stepping-stone in the evolutionary development of a jamming signal, a unique acoustic countermeasure in the long-standing arms race between bats and their insect prey.

## References and Notes

1. D. R. Griffin, *Anim. Behav.* **8**, 141 (1960).
2. H. U. Schnitzler, E. K. V. Kalko, *Bioscience* **51**, 557 (2001).
3. L. A. Miller, A. Surlykke, *Bioscience* **51**, 570 (2001).
4. K. D. Roeder, *J. Insect Physiol.* **10**, 529 (1964).
5. K. D. Roeder, *Anim. Behav.* **10**, 300 (1962).
6. A. D. Blest, *Zoologica* **49**, 161 (1964).
7. S. J. Weller, N. L. Jacobsen, W. E. Conner, *Biol. J. Linn. Soc. London* **68**, 557 (1999).
8. D. L. Bates, M. B. Fenton, *Can. J. Zool.* **68**, 49 (1990).
9. D. C. Dunning, K. D. Roeder, *Science* **147**, 173 (1965).
10. D. C. Dunning, *Z. Tierpsychol.* **25**, 129 (1968).
11. L. Acharya, M. B. Fenton, *Can. J. Zool.* **70**, 1292 (1992).
12. D. C. Dunning, L. Acharya, C. B. Merriman, L. Dal Ferro, *Can. J. Zool.* **70**, 2218 (1992).
13. N. I. Hristov, W. E. Conner, *Naturwissenschaften* **92**, 164 (2005).
14. A. Surlykke, L. A. Miller, *J. Comp. Physiol. [A]* **156**, 831 (1985).
15. J. H. Fullard, M. B. Fenton, J. A. Simmons, *Can. J. Zool.* **57**, 647 (1979).
16. J. H. Fullard, J. A. Simmons, P. A. Saillant, *J. Exp. Biol.* **194**, 285 (1994).
17. L. A. Miller, *J. Comp. Physiol. [A]* **168**, 571 (1991).
18. J. Tougaard, J. H. Casseday, E. Covey, *J. Comp. Physiol. [A]* **182**, 203 (1998).
19. J. Tougaard, L. A. Miller, J. A. Simmons, in *Echolocation in Bats and Dolphins*, J. A. Thomas, C. F. Moss, M. Vater, Eds. (Univ. of Chicago Press, Chicago, 2004), pp. 365–380.
20. J. R. Barber, W. E. Conner, *J. Exp. Biol.* **209**, 2637 (2006).
21. Materials and methods are available as supporting material on *Science* Online.
22. J. R. Barber, W. E. Conner, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9331 (2007).
23. M. Davis, in *Neural Mechanisms of Startle Behavior*, R. C. Eaton, Ed. (Plenum Press, New York, 1984), pp. 287–351.
24. N. M. Marples, T. J. Roper, D. G. Harper, *Oikos* **83**, 161 (1998).
25. J. M. Ratcliffe, J. H. Fullard, *J. Exp. Biol.* **208**, 4689 (2005).
26. S. A. Kick, J. A. Simmons, *Neuroscience* **4**, 2725 (1984).
27. W. W. Wilson, C. F. Moss, in *Echolocation in Bats and Dolphins*, J. A. Thomas, C. F. Moss, M. Vater, Eds. (Univ. of Chicago Press, Chicago, 2004), pp. 365–380.
28. We thank F. Insana and the staff of the Southwestern Research Station for assistance collecting *B. trigona*; M. Cullen for providing bat care and training; N. Hristov and J. Muday for technical assistance; and B. Chadwell for software. M. Conner, J.-N. Jasmin, T. Eisner, and the Wake Forest University ecology group provided editorial comments. Funding for research was provided by National Science Foundation Grant IOB-0615164 to W.E.C. and a W.F.U. Dean's Fellowship to A.J.C.
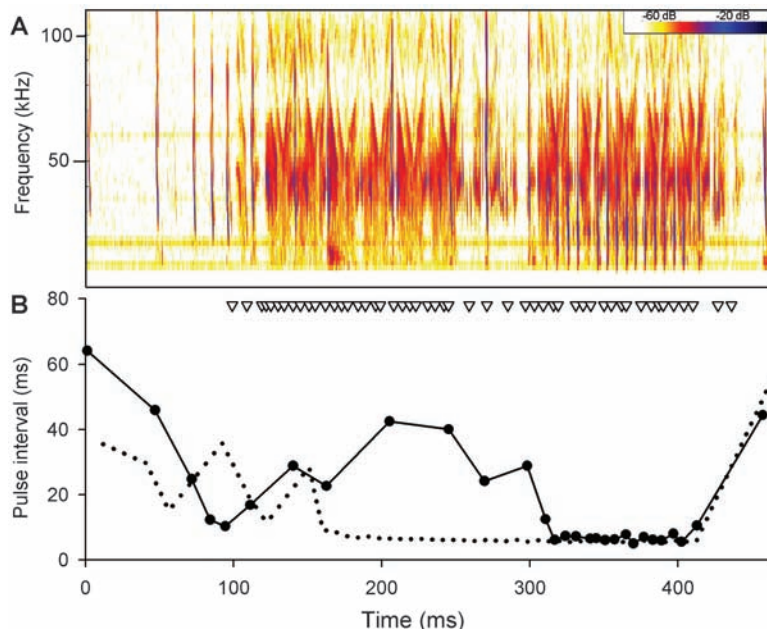
**Fig. 3.** Example echolocation attack sequence. (**A**) Spectrogram showing a big brown bat attacking a clicking *B. trigona*. (**B**) Pulse-interval graph of echolocation attack. Each circle connected by solid lines indicates the time of a sonar pulse and the duration of the following pulse interval. Open triangles indicate moth click half modulation cycles. The dotted line shows an attack on a control moth for reference.

**Table 1.** Contact rates of big brown bats on *B. trigona* and noctuid control moths.

| Bat | B. trigona | | Noctuid control | | $\chi^2$ | df | P |
|---|---|---|---|---|---|---|---|
| | n | Percent contact | n | Percent contact | | | |
| J1 | 19 | 16 | 28 | 93 | 28.45 | 1 | <0.001 |
| J2 | 15 | 20 | 28 | 100 | 27.22 | 1 | <0.001 |
| A1 | 25 | 20 | 28 | 86 | 23.02 | 1 | <0.001 |

# Functional Amyloids As Natural Storage of Peptide Hormones in Pituitary Secretory Granules

Samir K. Maji,[1]* Marilyn H. Perrin,[2] Michael R. Sawaya,[3] Sebastian Jessberger,[4] Krishna Vadodaria,[4] Robert A. Rissman,[5] Praful S. Singru,[6] K. Peter R. Nilsson,[7] Rozalyn Simon,[7] David Schubert,[8] David Eisenberg,[3] Jean Rivier,[2] Paul Sawchenko,[2] Wylie Vale,[2] Roland Riek[1,9]†

Amyloids are highly organized cross–β-sheet–rich protein or peptide aggregates that are associated with pathological conditions including Alzheimer's disease and type II diabetes. However, amyloids may also have a normal biological function, as demonstrated by fungal prions, which are involved in prion replication, and the amyloid protein Pmel17, which is involved in mammalian skin pigmentation. We found that peptide and protein hormones in secretory granules of the endocrine system are stored in an amyloid-like cross–β-sheet–rich conformation. Thus, functional amyloids in the pituitary and other organs can contribute to normal cell and tissue physiology.

Cells transport newly synthesized secretory proteins and peptides in vesicles via the endoplasmic reticulum (ER) and Golgi for release into the extracellular space (1, 2). Some secretory cells, such as neuroendocrine cells and exocrine cells, store secretory proteins and peptides for extended time periods in a highly concentrated form in membrane-enclosed electron-dense cores termed "secretory granules" (1, 3, 4), which are derived from the Golgi complex. The dense cores of these granules are made up of large, insoluble secretory protein and peptide aggregates that are formed by self-association (4–6). The granules are not amorphous but possess a distinct molecular organization, possibly of crystalline structures (7) or large intermolecular aggregates (5, 8).

Amyloid fibrils are cross–β-sheet structures that are primarily associated with several neurodegenerative diseases including Alzheimer's disease. However, amyloid fibril formation also provides biologically functional entities termed functional amyloids (9) that are present in *Escherichia coli* (10), silkworms (11), fungi (12), and mammalian skin (13). The cross–β-sheet motif is composed of intermolecular β sheets along the fibril axis with the β strands aligned perpendicularly to the fibril axis. An amyloid-like structure of peptide and protein hormones in secretory granules could explain most of their properties.

To address the question of whether peptide and protein hormones are stored in secretory granules in an amyloid-like aggregation state, we first asked whether a diverse set of peptide and protein hormones could form amyloids in vitro at granule-relevant pH 5.5. We randomly selected 42 peptide and protein hormones from multiple species and organs, some linear and some cyclic, with a variety of different three-dimensional structures (table S2). This set of hormones was assayed for a capacity to form amyloids by the amyloid-specific dyes thioflavin T (Thio T) and Congo Red (CR), by luminescent conjugated polyelectrolyte probes (LCPs), by the conformational transition into β-sheet–rich structure measured by circular dichroism (CD) spectroscopy, and by the presence of fibrils in electron microscopy (EM) images. Furthermore, x-ray fiber diffraction was measured for a subset of hormones (table S1). Only 10 of the 42 hormones showed considerable formation of amyloids (table S1 and figs. S1 and S2).

Given that only one-quarter of peptides spontaneously formed amyloid fibrils in vitro (table S1) and the possible involvement of glycosaminoglycans (GAGs) in the formation of both secretory granules and amyloid fibrils (14–22), the amyloid formation of all 42 peptides and proteins was monitored in the presence of low molecular weight (LMW) heparin as a representative of GAGs. After 2 weeks of incubation in the presence of heparin, most hormones formed amyloid fibrils (table S1) as assessed by EM (Fig. 1A and fig. S3), Thio T binding (fig. S1), CD (table S1 and fig. S4), CR binding (table S1 and figs. S5A and S5B), LCP binding (fig. S5C), and x-ray fiber diffraction (fig. S5D). In addition, the algorithm TANGO predicted 18 hormones to be aggregation-prone

(fig. S6). Thus, of the 42 hormones, 31 formed amyloid fibrils in the presence of heparin by all methods studied. Nonconclusive data were obtained for four hormones (23) and the α-helical corticotropin-releasing factor (α-CRF) served as a negative control (23). Possible explanations for the lack of amyloid aggregation of the remaining six hormones may be that these hormones do not form amyloid-like aggregates when stored in secretory granules, that fibrillization conditions were not optimal, and/or that another substance might be required for aggregation. Indeed, human prolactin did not form amyloids in the presence of heparin, but did in the presence of chondroitin sulfate A, which is a GAG found in prolactin-specific granules (16) (Fig. 1A, table S1, and fig. S1).

None of the adrenocorticotropic hormones (ACTHs) studied formed amyloid-like aggregates (table S1). Because ACTH and β-endorphin are processed from the same prohormone (pro-opiomelanocortin) and secreted together in a regulated secretory pathway, we hypothesized that ACTH might need the amyloid-forming β-endorphin as an aggregation partner for storage in secretory granules. An ACTH–β-endorphin 1:1 mixture in the presence of heparin formed amyloid fibrils (Fig. 1B) that comprised ACTH, as measured by dot plot (Fig. 1C) and Trp fluorescence (Fig. 1D). The Trp/Tyr measurements also indicate that the ACTH–β-endorphin amyloid fibrils are composed of a 1:2 ratio of ACTH and β-endorphin present either in mixed or individual fibrils (Fig. 1D). Because seeding experiments of ACTH with β-endorphin fibrils failed, we suggest that ACTH and β-endorphin form mixed fibrils. Because fibril-forming hypothalamic human CRF was not able to induce aggregation of the pituitary ACTH, the ACTH–β-endorphin aggregation appeared to be specific [fig. S8, D and E; see also (23) and fig. S7]. Coaggregation between β-endorphin and ACTH was further supported by a colocalization study of the two hormones in the tumor cell line AtT20 using double immunohistochemistry (Fig. 1E and fig. S7). A detailed analysis revealed that all the stained ACTH colocalized with β-endorphin, whereas some β-endorphin did not colocalize with ACTH, which supports the in vitro observation that ACTH aggregates only in the presence of β-endorphin. To show that the ACTH coaggregation with β-endorphin was not an isolated case, we document coaggregation of ghrelin with obestatin in (23) and fig. S8.

The concept that peptide and protein hormones are stored in secretory granules in an amyloid-like aggregation state is challenged by the notion that amyloid fibrils are very stable and do not release monomers [but see (21)], a prerequisite upon granule secretion. To demonstrate that amyloid fibrils formed by hormones are able to release monomers, we performed an amyloid release assay (21) for a selection of hormone fibrils both at granule-relevant pH 6 and at pH 7.4, at which the hormone aggregates

[1]Laboratory of Physical Chemistry, Eidgenössische Technische Hochschule (ETH) Zürich, Wolfgang-Paulistrasse 10, CH-8093 Zürich, Switzerland. [2]Clayton Foundation Laboratories for Peptide Biology, Salk Institute for Biological Studies, La Jolla, CA 92037, USA. [3]Howard Hughes Medical Institute, UCLA-DOE Institute for Genomics and Proteomics, Box 951570, University of California, Los Angeles, CA 90095, USA. [4]Institute of Cell Biology, Department of Biology, ETH Zürich, 8093 Zürich, Switzerland. [5]Department of Neuroscience, University of California, San Diego, La Jolla, CA 92093, USA. [6]Division of Endocrinology, Diabetes and Metabolism, Tufts Medical Center, Boston, MA 02111, USA. [7]Department of Chemistry, IFM, Linköping University, SE-581 83 Linköping, Sweden. [8]Cellular Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA. [9]Structural Biology, Salk Institute for Biological Studies, La Jolla, CA 92037, USA.

*Present address: School of Bioscience and Bioengineering, IIT Bombay, Powai, Mumbai 400076, India.

†To whom correspondence should be addressed. E-mail: roland.riek@phys.chem.ethz.ch

are exposed upon secretion (Fig. 2 and fig. S9). The monomer release was monitored outside the membrane by CD (Fig. 2B) and by ultraviolet absorption spectroscopy or Trp fluorescence spectroscopy (fig. S9), and the remaining aggregates (inside the dialysis membrane) were monitored by Thio T binding (Fig. 2A and fig. S9). All hormone fibrils studied appeared to release monomeric hormone upon dilution, because with time a drastic decrease of the Thio T binding was apparent and monomers were present outside the membrane. To show that the released monomers were functional, we investigated the monomer release of human CRF from its amyloid entity in more detail. The CD spectrum (Fig. 2B) indicated that the released human CRF comprises its functional helical structure. Furthermore, mono-

meric and aggregated human CRF activate cyclic adenosine monophosphate (cAMP) with similar efficacy in CHO cells expressing stably human type 1 CRF receptor (Fig. 2C), indicating that CRF amyloid fibrils release active monomer. The in vitro dialysis experiments with amyloid fibrils therefore resemble qualitatively the release abilities of isolated secretory granules (24).

Amyloid fibrils are thought to be toxic to neuronal cells (25). To test whether amyloids of hormones are toxic, we used the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide–formazan assay (MTT assay) with the neuronal cell line B12 (25). Upon addition of 20 μM Aβ(1-40) fibrils, MTT reduction was decreased by 40% (fig. S10A). In contrast, many of the hormone amyloids formed in the presence

of LMW heparin and all hormones that did not aggregate into amyloids induced no considerable decrease in MTT reduction at 20 μM. However, in contrast to monomers, amyloids of ovine CRF, mUcnIII, hUcnIII, human growth hormone–releasing factor, and GLP-2 showed a response similar to that of Aβ(1-40), whereas glucagon amyloids interfered with the reduction of MTT almost twice as much (26). To further support the observation that some amyloids from hormones are moderately toxic, we measured the neuronal cell survival of E18 rat hippocampal neurons after the addition of amyloid fibrils of Aβ(1-42) and human CRF, which showed a moderate effect in the MTT assay (fig. S10, B and C). Although Aβ(1-42) fibrils appeared to be one order of magnitude more
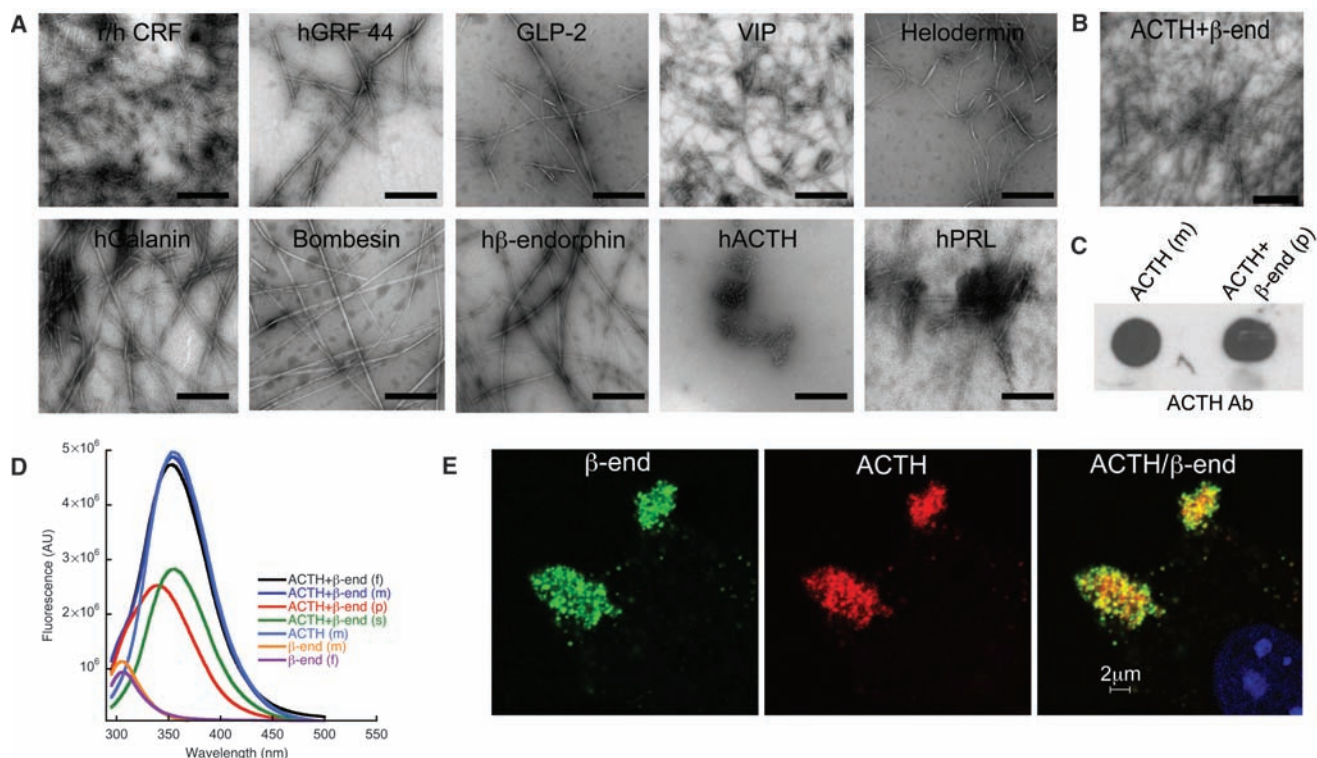


**Fig. 1.** Amyloid-like aggregation and coaggregation of hormones. (**A**) EM images of 10 hormones incubated for 14 days, indicating the formation of amyloid fibrils. The entire set of 42 hormones studied is shown in fig. S3. The aggregations of the hormones were followed at 37°C at a concentration of 2 mg/ml in the presence of 0.4 mM LMW heparin in 5% D-mannitol (pH 5.5) under slight agitation. Human prolactin (hPRL) was fibrillized in the presence of 400 μM chondroitin sulfate A. Transmission electron microscopy (TEM) of negative stained samples was performed. Scale bars, 500 nm. (**B** to **E**) Coaggregation of ACTH with β-endorphin measured by (B) EM, (C) dot blot, (D) Trp and Tyr fluorescence, and (E) colocalization in AtT20 cells. (B) EM of ACTH–β-endorphin mixture at 37°C at hormone concentrations of 1 mg/ml each in the presence of 0.4 mM LMW heparin in 5% D-mannitol (pH 5.5) incubated under slight agitation for 14 days. (C) The presence of ACTH in these amyloid fibrils of the ACTH–β-endorphin mixture was identified by dot blot staining with ACTH antibody of the aggregates harvested by centrifugation (p). For a positive control of the antibody staining, ACTH monomer (m) was used. (D) Trp and Tyr fluorescence was measured for an aggregated sample of ACTH–β-endorphin mixture, labeled with (f) and shown in black. High-speed centrifugation at 120,000g for 1 hour was performed with this sample, and the Trp and Tyr fluorescence was measured for

the corresponding pellet (p) in red, as well as the supernatant (s) in green. Furthermore, a fresh solution of a mixture of ACTH–β-endorphin (m) in blue, a fresh solution of ACTH (m) in cyan, a fresh solution of β-endorphin (m) in yellow, and an aggregated β-endorphin fibrillar sample (f) in violet were measured accordingly. The Trp fluorescence signal in pellet suggests that ACTH is present in fibrils because only ACTH has a Trp. The λ_max of the Trp signal is blue-shifted by 15 nm when compared to monomeric/ nonfibrillated ACTH, which suggests that the Trp is less solvent-exposed in the fibril structure than in its monomeric state. The fluorescence study suggests the aggregation of about half of the ACTH, because the Trp fluorescence intensities of the pellet containing fibrils and the corresponding supernatant containing ACTH monomer were similar. Because the Tyr fluorescence of monomeric β-endorphin, the β-endorphin fibrils harvested by centrifugation, and the harvested ACTH–β-endorphin fibrils were similar in intensity and no Tyr fluorescence signal was observed in the corresponding supernatant, the entire β-endorphin population appeared to have aggregated into amyloid fibrils. (E) Colocalization of the two hormones in the tumor cell line AtT20 by double immunohistochemistry with mouse ACTH (red) and rabbit endorphin antibodies (green). The nuclear marker 4´,6-diamidino-2-phenylindole is shown in blue.

toxic than human CRF fibrils, human CRF amyloids also influenced the survival of neurons with a median inhibitory concentration of ~20 μM. Although some hormone amyloids appear to be moderately toxic, their actual toxicity might be substantially diminished in vivo by their membrane-encapsulated state in the granules.

The above biophysical analysis of hormones supports the hypothesis that hormones are stored as amyloids in secretory granules. Below we offer direct evidence that secretory granules of the mouse pituitary tumor neuroendocrine cell line AtT20 are composed of amyloids. The AtT20

cell line synthesizes precursors to corticotropin (i.e., ACTH and β-endorphin) and correctly glycosylates, cleaves, and stores them in secretory granules (27). Secretory granules of AtT20 cells with and without membrane were purified (fig. S11) and showed typical ACTH-containing electron-dense granules with a diameter of 200 to 600 nm surrounded by membrane (fig. S11, B to D). The purified granules from AtT20 cells were amyloid-like in nature, as determined by amyloid-specific antibody (Fig. 3A), Thio T binding (Fig. 3B), CR binding (fig. S12), and CR birefringence (Fig. 3C). Finally, x-ray fiber diffraction was

measured of purified membraneless secretory granules (fig. S11B). A near–membrane-free preparation of granules was used because membrane lipids contribute a very strong reflection at 4.1 Å, close to the 4.7 Å reflection expected for a cross–β-sheet structure. The major reflections observed were at 4.7 Å, interpreted as the spacing between strands in a β sheet, and a diffuse reflection at ~10 Å, interpreted as the spacing between β sheets (Fig. 3D). These reflections are typically observed for amyloid-like fibrils. The circular profiles of these reflections, rather than the normal orthogonal positions for the two reflections,



**Fig. 2.** Release of monomeric, α-helical, and functional CRF from its amyloid fibrils. CRF amyloid fibrils were dialyzed against buffer with a 10-kD cutoff membrane. (**A**) Time-dependent decrease of Thio T fluorescence inside the membrane at two pHs as labeled. The decrease of Thio T indicates loss of amyloid fibrils due to dialysis. (**B**) Time-resolved CD spectroscopy outside the membrane measuring released CRF. The time-dependent increase of the signal indicates release of CRF from the amyloid. The released CRF is likely to be monomeric because of the 10-kD cut-off of the dialysis membrane. The CD spectrum of the released CRF is of helical structure, which corresponds to the active conformation of CRF. (**C**) Functional studies of monomeric and amyloid fibrillar sample of CRF by measuring in a hormone concentration–dependent manner the activation of intracellular cAMP in CHO cells stably expressing CRF-R1. Both samples show similar potencies.
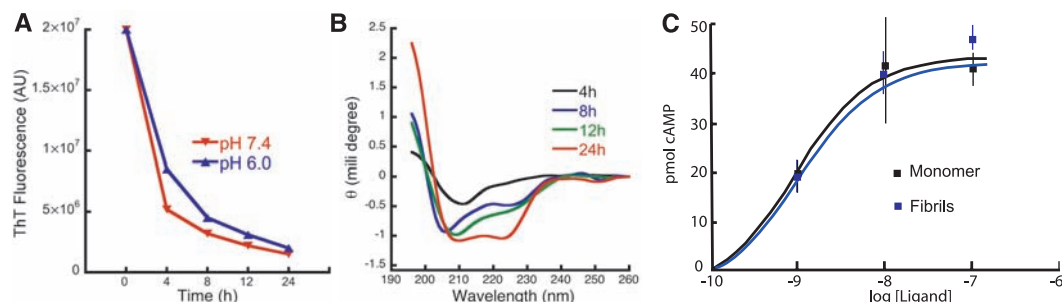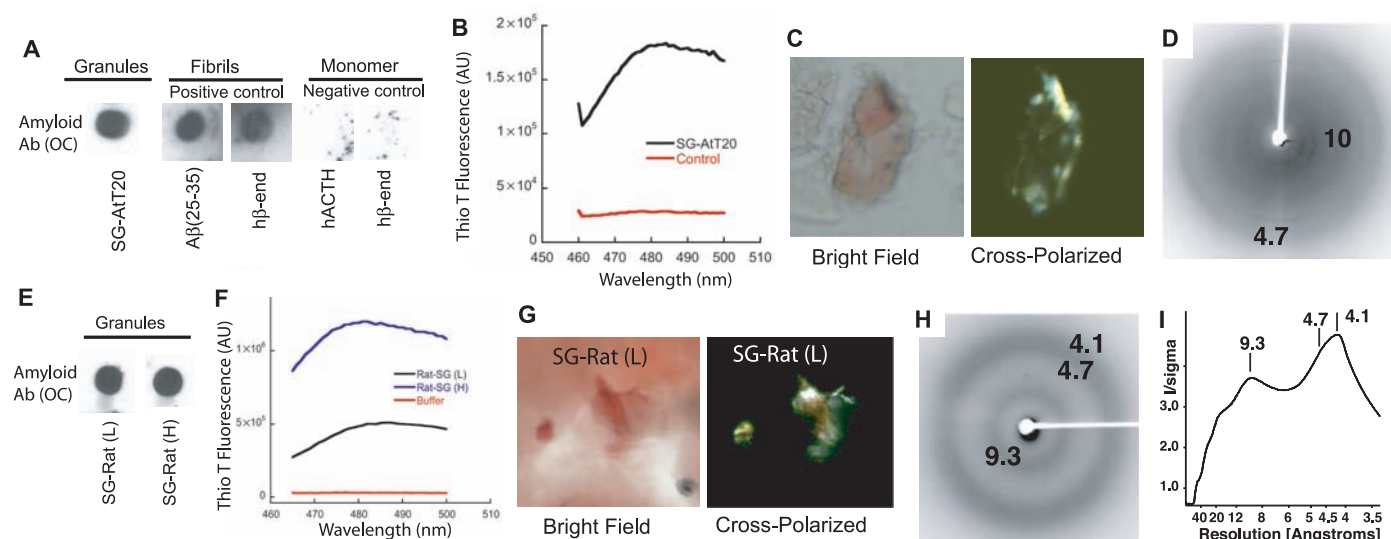


**Fig. 3.** Purified secretory granules from the AtT20 cell line and from rat pituitary are composed of an amyloid-like structure as determined by an amyloid-specific antibody OC (**A** and **E**), the amyloid-specific dye Thio T (**B** and **F**), CR birefringence (**C** and **G**), and x-ray fiber diffraction (**D**, **H**, and **I**). (A) Dot blot staining of purified secretory granules from AtT20 cells against the amyloid-specific antibody OC. As positive controls, stains of amyloid fibrils of Aβ(25-35) fibrils and β-endorphin fibrils (in presence of LMW heparin) are shown. As negative controls, stains of monomeric human ACTH and β-endorphin are shown. (B) Thio T fluorescence between 460 and 500 nm is shown for purified secretory granules (black) and buffer only (red). Thio T was excited at 450 nm. The strong binding of Thio T to secretory granules suggests that they are composed of amyloid-like structures. (C) Congo Red birefringence of purified secretory granules. CR birefringence is shown that suggests the presence of ordered (amyloid) aggregates. The picture

represents the bright-field microscope image with 10× resolution. The same section is shown under cross-polarized light with 10× magnification, respectively. (D) X-ray fiber diffraction of purified membrane-depleted secretory granules that were treated with 1% Lubrol PX. The two reflections at 4.7 Å and ~10 Å consistent with cross–β-sheet structure are labeled. [(E) to (I)] Purified light (L) and heavy (H) secretory granules from rat pituitary are composed of amyloid-like structure as assessed by an amyloid-specific antibody OC (E), the amyloid-specific dye Thio T (F), CR birefringence (G), and x-ray fiber diffraction [(H) and (I)]. The same measurement parameters as in [(A) to (D)] were used. In (I) the x-ray fiber diffraction measurements of purified membraneless secretory granules are shown as a full image radial profile. The two reflections at ~4.7 Å and ~9.3 Å consistent with cross–β-sheet structure are labeled. In addition, a strong reflection at 4.1 Å is present, attributed to the remaining lipid content of the granules under study.

show that the amyloid-like entities in granules are not strongly aligned.

Similarly, secretory granules of type light (L) and heavy (H) purified from rat pituitary (fig. S13) also contained amyloid-like material, as evidenced by amyloid-specific antibody binding (Fig. 3E), Thio T binding (Fig. 3F), CR binding (fig. S12, C and D), and birefringence (Fig. 3G) (23). Furthermore, the x-ray diffraction pattern with membraneless granules of type L show the typical cross–β-sheet reflections at 4.7 Å and at ~10 Å, in addition to a 4.1 Å reflection attributed to remaining membrane (Fig. 3, H and I). The circular profiles for these reflections indicate that the amyloid-like entities in granules are not strongly aligned.

Immunohistochemical studies were performed with mouse pituitary tissue to probe whether secretory granules in neuroendocrine tissues are amyloid-like in nature. The positive binding of the amyloid-specific dye Thio S both in the interior and posterior lobes of the pituitary was indicative of the abundant presence of amyloids in the mouse pituitary (Fig. 4). To examine whether the Thio S–detected amyloids were composed of endocrine hormones, we measured colocalization of Thio S with hormone-specific antibodies. There was almost complete colocalization between Thio S and the hormones ACTH, β-endorphin, prolactin, and growth hormone (GH) in the interior lobe, ACTH in the intermediate lobe, and oxytocin and vasopressin in the posterior lobe of the pituitary, respectively (Fig. 4), strongly supporting their storage in the secretory granules to be extensively amyloid-like. Similar results were obtained for a colocalization study between the fibril-specific antibody OC and the corresponding hormone antibodies.

The proposed amyloid-like conformation of peptide/protein hormones in secretory granules may explain the processes of granule formation including hormone selection, membrane surrounding and inert hormone storage, and subsequently the release of hormones from the granules (fig. S14). We propose that in the Golgi, amyloid aggregation of the (pro)hormone is initiated spontaneously above a critical (pro)hormone concentration and/or in the presence of helper molecules such as GAGs in parallel to a possible prohormone processing. Alternatively, because the prohormone may aggregate less into an amyloid entity than its hormone counterpart (28), the prohormone processing at critical hormone concentrations may initiate the aggregation. Because the formation of amyloid fibrils is amino acid sequence–specific, the initiated amyloid aggre-gation of the (pro)hormone is selective, yielding granule cores composed of specific hormones only. Specific coaggregation of some hormones may be possible because some amyloid proteins are able to cross-seed (29) (Fig. 1). The amyloid aggregation thereby sorts the protein/peptide hormones into secretory granule cores, concentrates them to the highest density possible, and excludes non–aggregation-prone, constitutively secreted proteins. During the aggregation process, the hormone amyloids become surrounded by membrane, separate from the Golgi, and form mature granules. The membrane attraction may be spontaneous because membrane binding seems to be an inherent property of amyloid aggregates (30). Alternatively, because the cross–β-sheet represents a single structural epitope, it may serve as a possible recognition motif of an unknown granule-recruiting membrane protein. Once the secretory granules are formed, they can be stored for long durations because the amyloid entity provides a very stable depot. Upon signaling, secretory granules are secreted and the cross–β-sheet structure of the amyloid enables a controlled release of monomeric, functional hormone (21), which might be supported by chaperones.

The functional amyloid state of many endocrine hormones in secretory granules of the pi-



**Fig. 4.** Immunohistochemical staining of the mouse pituitary with Thio S (green) and polyclonal antibodies (red) to (**A**) prolactin, (**B**) and (**C**) ACTH, (**D**) growth hormone (GH), (**E**) oxytocin, and (**F**) vasopressin. Regions of the pituitary selected are anterior lobe [(A), (B), and (D)], intermediate lobe (C), and posterior lobe [(E) and (F)]. Scale bars, 20 μm.
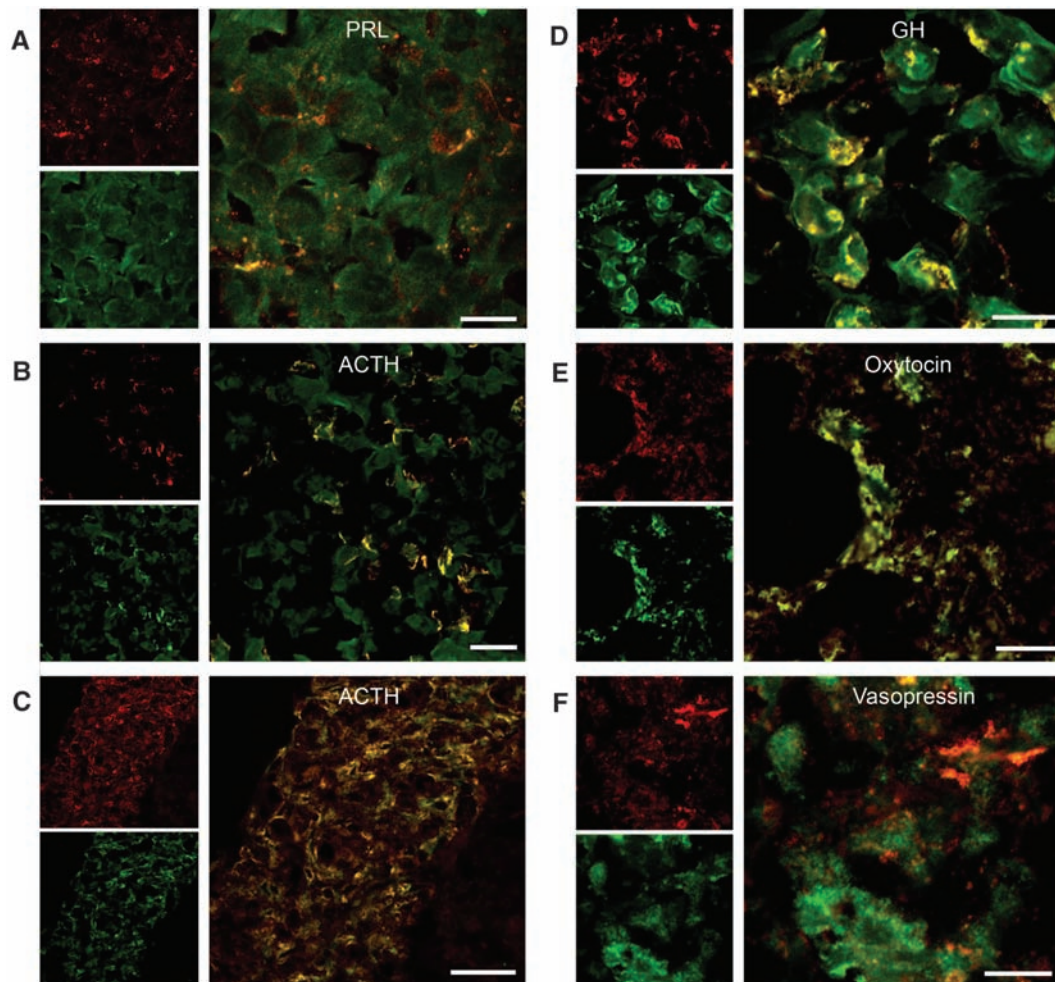
tuitary (Fig. 4), and possibly the hypothalamus (e.g., CRF) and pancreas (e.g., somatostatin), contrasts directly the historical disease association of amyloids both in the brain (e.g., tau, α-synuclein, Aβ, and prion protein) and in the pancreas (e.g., insulin and amylin). On the one hand, this challenges once more the "amyloid hypothesis" that amyloids are the most toxic culprit in amyloid diseases (31). On the other hand, hormone amyloids may not be (very) toxic because the hormone amyloids are stored inside the granules and the amyloid aggregation of hormones for secretory granule formation may be highly regulated. This regulation may include the processing of prohormones that aggregate more slowly than the hormone counterpart (28) or the required presence of helper molecules to induce aggregation, as demonstrated for prolactin (note that prolactin lacks a prohormone). Furthermore, the hormone amyloids are stored in an "inert" membrane container and the amyloid fibrils dissociate upon secretion. Such regulation requires a functional protein homeostasis. If the protein homeostasis is altered under certain conditions such as diet, stress, or age, hormone aggregation may be out of control and disease-associated amyloid aggregation of hormones may occur (see below). Whether such aggregations cause disease, or are an indirect effect of the protein homeostasis altered by disease, remains to be determined.

There are many associations between amyloid processing and function of secretory granules: (i) Aggregation of hormones into secretory granules is an intracellular process, and upon secretion not all of the aggregates dissolve completely (32). Similarly, Aβ may be present in granules (33), and Aβ aggregation may be initiated intracellularly and end up as a pathological hallmark in the extracellular space (31). (ii) The malaria drug chloroquine interferes with the formation of secretory granules as well as prion infectivity (34, 35). (iii) The natural compound curcumin interferes with both the progression of Alzheimer's disease (36) and the release of endocrine hormones (37). (iv) Several endocrine hormones that may be stored in secretory granules in an amyloid-like state are present as amyloids in amyloid diseases. These include amylin associated with diabetes type II, calcitonin associated with medullary carcinoma of the thyroid, and atrial natriuretic factor with atrial amyloidosis (38). Hence, the presence of many functional amyloids in the body, together with the apparent tight link between functional and disease-associated amyloids in their processing, biophysical, and biochemical properties, requires a rethinking of the relationship between aggregation and function of polypeptides, and the correlation between amyloid aggregation and toxicity.

**References and Notes**
1. R. B. Kelly, Science 230, 25 (1985).
2. R. B. Kelly, Nature 326, 14 (1987).
3. P. Arvan, D. Castle, Biochem. J. 332, 593 (1998).
4. P. S. Dannies, Mol. Cell. Endocrinol. 177, 87 (2001).
5. P. Arvan, B. Y. Zhang, L. Feng, M. Liu, R. Kuliawat, Curr. Opin. Cell Biol. 14, 448 (2002).
6. S. A. Tooze, Biochim. Biophys. Acta 1404, 231 (1998).
7. F. Miller, E. de Harven, G. E. Palade, J. Cell Biol. 31, 349 (1966).
8. C. Keeler, M. E. Hodsdon, P. S. Dannies, J. Mol. Neurosci. 22, 43 (2004).
9. D. M. Fowler, A. V. Koulov, W. E. Balch, J. W. Kelly, Trends Biochem. Sci. 32, 217 (2007).
10. M. M. Barnhart, M. R. Chapman, Annu. Rev. Microbiol. 60, 131 (2006).
11. V. A. Iconomidou, G. Vriend, S. J. Hamodrakas, FEBS Lett. 479, 141 (2000).
12. M. L. Maddelein, S. Dos Reis, S. Duvezin-Caubet, B. Coulary-Salin, S. J. Saupe, Proc. Natl. Acad. Sci. U.S.A. 99, 7402 (2002).
13. D. M. Fowler et al., PLoS Biol. 4, e6 (2006).
14. S. O. Kolset, K. Prydz, G. Pejler, Biochem. J. 379, 217 (2004).
15. H. A. Reggio, G. E. Palade, J. Cell Biol. 77, 288 (1978).
16. A. Zanini et al., J. Cell Biol. 86, 260 (1980).
17. P. Rosa, A. Zanini, Eur. J. Cell Biol. 31, 94 (1983).
18. S. S. Carlson, R. B. Kelly, J. Biol. Chem. 258, 11082 (1983).
19. R. Kisilevsky, J. Struct. Biol. 130, 99 (2000).
20. J. A. Cohlberg, J. Li, V. N. Uversky, A. L. Fink, Biochemistry 41, 1502 (2002).
21. S. K. Maji et al., PLoS Biol. 6, e17 (2008).
22. J. Y. Suk, F. Zhang, W. E. Balch, R. J. Linhardt, J. W. Kelly, Biochemistry 45, 2234 (2006).
23. See supporting material on Science Online.
24. G. Giannattasio, A. Zanini, J. Meldolesi, J. Cell Biol. 64, 246 (1975).
25. Y. Liu, D. Schubert, J. Neurochem. 71, 2322 (1998).
26. S. Onoue et al., Pharm. Res. 21, 1274 (2004).
27. B. Gumbiner, R. B. Kelly, Proc. Natl. Acad. Sci. U.S.A. 78, 318 (1981).
28. I. T. Yonemoto, G. J. Kroon, H. J. Dyson, W. E. Balch, J. W. Kelly, Biochemistry 47, 9900 (2008).
29. B. I. Giasson et al., Science 300, 636 (2003).
30. G. P. Gellermann et al., Proc. Natl. Acad. Sci. U.S.A. 102, 6297 (2005).
31. J. A. Hardy, G. A. Higgins, Science 256, 184 (1992).
32. M. G. Farquhar, in Subcellular Structure and Function in Endocrine Organs, H. Heller, K. Lederis, Eds. (Cambridge Univ. Press, Cambridge, 1971), pp. 79–122.
33. V. Hook, I. Schechter, H. U. Demuth, G. Hook, Biol. Chem. 389, 993 (2008).
34. H. P. Moore, B. Gumbiner, R. B. Kelly, Nature 302, 434 (1983).
35. C. Korth, B. C. May, F. E. Cohen, S. B. Prusiner, Proc. Natl. Acad. Sci. U.S.A. 98, 9836 (2001).
36. F. Yang et al., J. Biol. Chem. 280, 5892 (2005).
37. M. Miller, S. Chen, J. Woodliff, S. Kansra, Endocrinology 149, 4158 (2008).
38. F. Chiti, C. M. Dobson, Annu. Rev. Biochem. 75, 333 (2006).
39. We thank S. Maji for the cartoon drawing in fig. S14; S. Pisani and C. Peto for technical assistance; C. R. Grace and C. Tzitzilonis for valuable suggestions; P. S. Dannies and M. E. Hodsdon for valuable suggestions and the plasmid of human prolactin; and C. Glabe for the OC antibody. Supported by the Swiss National Foundation, Clayton Medical Research Foundation Inc., and National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) grant PO1 DK026741-29. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIDDK or NIH. W.V. is a cofounder, consultant, equity holder, and member of the Board of Directors and Scientific Advisory Board of Neurocrine Biosciences Inc. and Acceleron Pharma Inc. J.R. is founder and owner of Sentia Medical Sciences Inc. and is coinventor in several patents covering CRF receptor ligands.

# RIP3, an Energy Metabolism Regulator That Switches TNF-Induced Cell Death from Apoptosis to Necrosis

Duan-Wu Zhang,[1] Jing Shao,[1] Juan Lin,[1] Na Zhang,[1] Bao-Ju Lu,[2] Sheng-Cai Lin,[1] Meng-Qiu Dong,[2] Jiahuai Han[1]*

Necrosis can be induced by stimulating death receptors with tumor necrosis factor (TNF) or other agonists; however, the underlying mechanism differentiating necrosis from apoptosis is largely unknown. We identified the protein kinase receptor-interacting protein 3 (RIP3) as a molecular switch between TNF-induced apoptosis and necrosis in NIH 3T3 cells and found that RIP3 was required for necrosis in other cells. RIP3 did not affect RIP1-mediated apoptosis but was required for RIP1-mediated necrosis and the enhancement of necrosis by the caspase inhibitor zVAD. By activating key enzymes of metabolic pathways, RIP3 regulates TNF-induced reactive oxygen species production, which partially accounts for RIP3's ability to promote necrosis. Our data suggest that modulation of energy metabolism in response to death stimuli has an important role in the choice between apoptosis and necrosis.

Cell death occurs through morphologically distinct processes of apoptosis and necrosis (1). Some necrosis is regulated, via pathways differing from those controlling classical apoptosis, although necrosis/apoptosis interconnectivity has been observed (1–4). Caspase inhibition, which distinguishes apoptotic and non-apoptotic cell death, sometimes shifts apoptosis to necrosis or enhances necrosis (5–8). Receptors containing death domains can induce a form of regulated necrosis through kinase activity of RIP1 (receptor-interacting protein 1) (3, 5, 6). Mitochondrial generation of reactive oxygen species (ROS) is essential for this type of necrosis (9). A genome-wide small interfering RNA (siRNA) screen has identified a number of genes involved in the signaling network of death domain receptor–mediated necrosis (10), but the precise mechanisms underlying programmed necrosis and the apoptosis/necrosis molecular switch remain unclear.

Although NIH 3T3 cells typically undergo apoptosis in response to tumor necrosis factor (TNF) stimulation (10), caspase-independent cell death has been reported in one NIH 3T3 line (termed N cells here) (11). We compared the caspase-dependence of TNF-induced cell death between NIH 3T3 cells obtained from American Type Culture Collection (termed A cells here) and N cells. Benzyloxycarbonyl-Val-Ala-Asp-fluoromethylketone (zVAD) inhibited TNF-induced cell death in A cells (Fig. 1A) but enhanced it in N cells (Fig. 1B). TNF- or TNF+zVAD–treated N cells had necrotic morphologies. Because both cell lines were derived from the same cell population, gene expression changes in the N cells could explain the switch from apoptosis to necrosis in the NIH 3T3 cells. Affymetrix microarray analysis revealed that a few genes that have the potential to regulate cell death or TNF signaling were expressed differently in A and N cells (table S1). The expression of RIP3, Calpain 5, Serpinb1a, and Serpinb9b increased in N cells, whereas expression of Calpain 6, Aifm1, and Serpine2 decreased. TNF or zVAD treatment had no effect on the expression of these genes in either A or N cells. We decreased expression of the increased genes and increased expression of the decreased genes in N cells, then analyzed TNF-induced cell death and zVAD enhancement of TNF-induced cell death. Only the siRNA-knockdown of RIP3 in N cells inhibited zVAD's enhancement of cell death (fig. S1). We therefore focused on RIP3.

RIP3 is a protein kinase that has an N-terminal kinase domain similar to that found in RIP1, RIP2, and RIP4, but its C-terminal domain has no sequence similarity to any known protein domains (12, 13). Transient expression of RIP3 in some cell lines causes cell death (12). Cleavage of RIP3, RIP1, and RIP4 by caspases was observed after activation of death receptors (8, 13, 14). Modulating nuclear factor κB (NF-κB) activation by means of RIP3 overexpression was observed (15, 16); however, deletion of RIP3 caused no signs of deregulated NF-κB signaling (16). The biological function of RIP3 is thus largely unknown.

Differing expression of RIP3 in A and N cells was confirmed by means of Western blotting (Fig. 1C) (17). To study the role of RIP3, we used lentiviral vectors to deliver short hairpin RNA (shRNA) or cDNA into A, N, or other cells because lentivirus can infect with 100% efficiency. RIP3 knockdown by means of siRNA in N cells appeared to switch the type of cell death to that of A cells because TNF-induced cell death in RIP3-knockdown N cells was not enhanced but rather inhibited by zVAD (Fig. 1D). A lacZ-targeting siRNA (Ctrl) showed no effect on the type of cell death in N cells (Fig. 1D). RIP3 expression in A cells switched TNF-induced apoptosis to caspase-

independent cell death (Fig. 1E). zVAD did not inhibit but rather enhanced cell death (Fig. 1E). RIP3 kinase activity was required to change the type of cell death because a catalytically inactive RIP3 mutant (RIP3-K51A) did not cause A cell death to become caspase-independent (fig. S2). Deletion and point mutations revealed that the RIP1-interacting RHIM (RIP homotypic interaction motif) domain in RIP3 is required for necrosis (fig. S2).

Although zVAD may also participate in the induction of cell death in some cell lines (10, 18), we established that RIP3 has a role in necrosis in the absence of zVAD because it is required for TNF-induced necrosis in L929 cells (fig. S3A). We then analyzed whether RIP3 is required for

zVAD-mediated necrosis. zVAD can trigger necrosis of lipopolysaccharides (LPS)– or TNF-stimulated primary peritoneal macrophages (19) but failed to promote cell death of RIP3$^{-/-}$ macrophages stimulated with TNF or LPS (Fig. 1F). It appears that RIP3 functions in necrosis induced by physiological stimuli and is required for zVAD to induce or enhance necrosis. Thus, RIP3 is essential for some forms of necrosis.

RIP1 is required for death receptor agonists to activate necrosis (5, 6, 8, 18). Although we detected no difference in RIP1 expression in A and N cells, depletion of RIP1 with shRNA in N cells inhibited TNF+zVAD–induced necrosis (Fig. 2A). Similarly, depletion of RIP1 from L929 cells also blocked TNF-induced necrosis (fig. S3B).
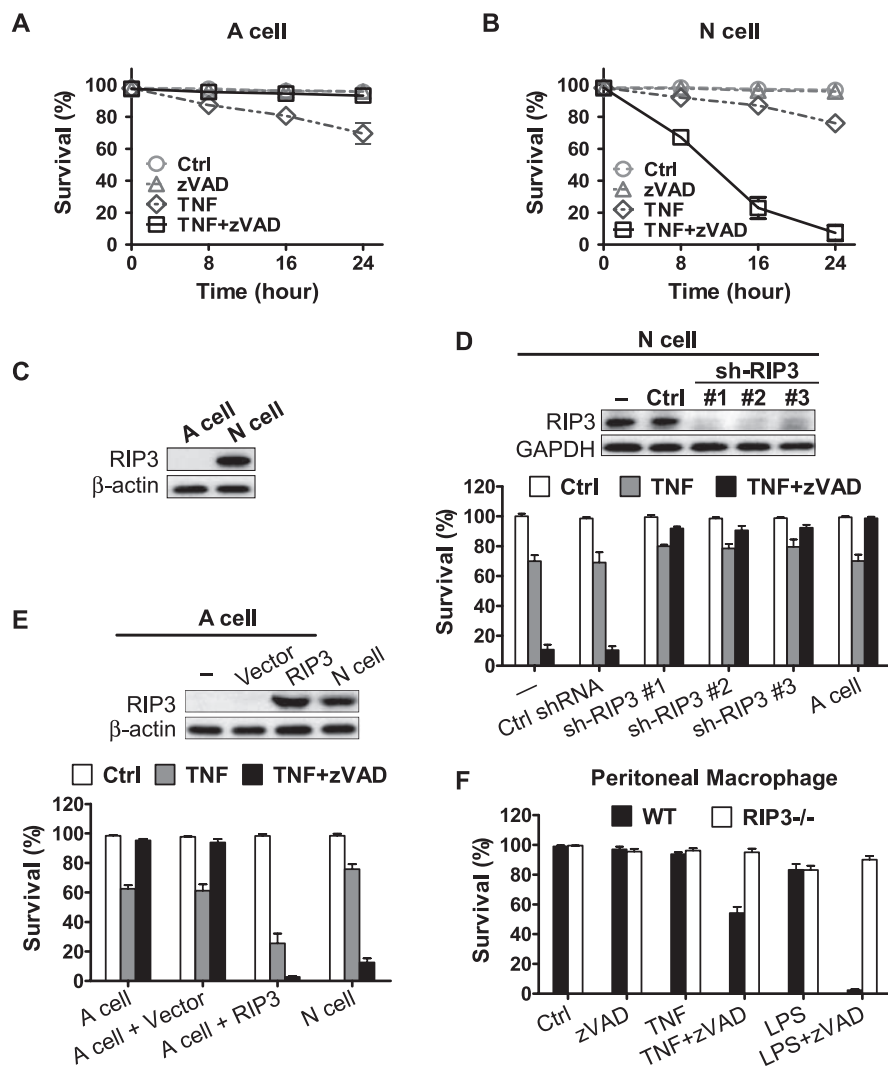


**Fig. 1.** Effects of RIP3 on type of cell death. (**A**) Viabilities of A cells treated with medium (Ctrl), zVAD (20 μM), TNF (30 ng/ml), or TNF+zVAD. (**B**) Viabilities of N cells treated as described in (A). (**C**) Immunoblotting analysis with antibody to RIP3. (**D**) Viabilities of N cells treated with or without (−) a lentiviral vector encoding LacZ shRNA (Ctrl) or RIP3-shRNAs for 48 hours then stimulated with or without TNF or TNF+zVAD for 24 hours. RIP3 protein amounts in these cells were analyzed by means of immunoblotting 48 hours after infection. (**E**) Viabilities of A cells infected with a lentivirus encoding nothing (vector) or RIP3 for 36 hours then stimulated with or without TNF or TNF+zVAD for 24 hours. RIP3 protein amounts were analyzed 36 hours after infection. (**F**) Viabilities of wild-type and RIP3$^{-/-}$ mouse peritoneal macrophages treated with medium (Ctrl), zVAD, TNF, TNF+zVAD, LPS (100 ng/ml), or LPS+zVAD for 24 hours. Data are the mean ± SD of triplicates [(A), (B), and (D) to (F)] and are representative of two to five experiments [(A) to (F)].

[1]Key Laboratory of the Ministry of Education for Cell Biology and Tumor Cell Engineering, School of Life Sciences, Xiamen University, Xiamen, Fujian 361005, China. [2]National Institute of Biological Sciences, Beijing 102206, China.

*To whom correspondence should be addressed. E-mail: jhan@xmu.edu.cn

Therefore, both RIP1 and RIP3 are required for TNF-induced and zVAD-enhanced necrosis.

Transient RIP1 expression caused cell death in both wild-type and RIP3$^{-/-}$ mouse embryo fibroblast (MEF) cells (Fig. 2B). zVAD inhibited RIP1-induced cell death in RIP3$^{-/-}$ cells but not in wild-type cells, so RIP3 may have a role in determining the type of RIP1-induced cell death. Consistently, RIP3-induced cell death was not inhibited by zVAD. Analyzing RIP1 and RIP3 expression in the knockout cell lines revealed that the previously described RIP1$^{-/-}$ MEF line (6) is defective in both RIP1 and RIP3 expression, and we therefore named it RIP1$^{-/-}$RIP3$^{d}$ (Fig. 2B, top). We reconstituted RIP3 expression in this cell line and named it RIP1$^{-/-}$*. RIP1- and RIP3-overexpression–induced death in RIP1$^{-/-}$* cells was similar to that observed in wild-type MEF cells, and cell death in RIP1$^{-/-}$RIP3$^{d}$ cells was similar to that in RIP3$^{-/-}$ cells (Fig. 2B), supporting the idea that RIP1 induces apoptosis in the absence of RIP3 and induces necrosis when RIP3 is present.

RIP1 overexpression induced apoptosis in A cells and necrosis in RIP3-rich N cells, whereas RIP3 overexpression induced necrosis in both A and N cells (Fig. 2C). Unlike MEF cells, necrosis in these cells was enhanced by zVAD, but only when RIP3 was present (Fig. 2C). We then found that zVAD did not enhance RIP1-induced N cell death when RIP3 was depleted (Fig. 2D). Thus, RIP3 is required for zVAD to enhance necrosis.

To find RIP3 targets, we identified interacting proteins during necrosis. Using a lentiviral vector, we expressed flag-tagged RIP3 in A cells and confirmed that flag-RIP3 changed the death type of A cells to that of N cells. Flag-RIP3 was then purified in a complex with its associated proteins from A cells treated with nothing or TNF+zVAD for 30 or 120 min. Protein constituents were identified by means of liquid chromatography–tandem mass spectrometry (LC-MS/MS). The raw data set containing nonredundant proteins (table S2) indicated dynamic changes in the RIP3 complex because proteins identified in TNF+zVAD–treated samples were different from the untreated sample. Seven metabolic enzymes—glycogen phosphorylase (PYGL), glutamate-ammonia ligase (GLUL), glutamate dehydrogenase 1 (GLUD1), fructose-1,6-bisphosphatase 2 (FBP2), fumarate hydratase (FH), glycosyltransferase 25 domain containing 1 (GLT25D1), and isocitrate dehydrogenase 1 (IDH1)—were identified in the RIP3 complex in TNF+zVAD–treated cells. PYGL, GLUL, and GLUD1 were confirmed to interact with RIP3 by means of coimmunoprecipitation of coexpressed proteins in 293T cells (fig. S4).

PYGL catalyzes the rate-limiting step in the degradation of glycogen by releasing glucose-1-phosphate, thereby having a key role in using reserved glycogen as an energy source. We sought to determine whether RIP3 and PYGL interact during necrosis, but were unable to obtain an antibody to detect endogenous PYGL in N cells. We therefore established an N cell line stably expressing hemagglutinin (HA)–PYGL. Immuno-

precipitation using antibodies to RIP3 revealed that TNF+zVAD treatment for 120 min induced interaction of RIP3 and HA-PYGL (Fig. 3A). RIP3 and PYGL interaction was induced by TNF but not zVAD (Fig. 3B). We incubated flag-PYGL (purified from 293T cells) with or without glutathione S-transferase (GST)–RIP3 or GST-RIP3-K51A (purified from Escherichia coli) in a kinase buffer and found that RIP3, but not RIP3-K51A, enhanced PYGL activity in vitro (Fig. 3C). Because RIP3 may directly target one or more proteins copurified with flag-PYGL, we currently can only conclude that RIP3 directly or indirectly activates PYGL in vitro. By measuring PYGL activity in N cells and RIP3- or RIP1-knockdown cells before and after TNF or

TNF+zVAD treatment, we found that RIP3 and RIP1 are required for TNF- and TNF+zVAD–increased PYGL activity (Fig. 3D and fig. S5). The role of RIP3 in regulating PYGL was confirmed by use of RIP3$^{-/-}$ peritoneal macrophages (Fig. 3E). Additionally, expression of RIP3 with PYGL in 293T cells enhanced PYGL activity (fig. S6). To determine whether PYGL is involved in necrosis, we used siRNAs to deplete PYGL in N cells and found that reduction of PYGL partially inhibited TNF+zVAD–induced cell death (Fig. 3F). Thus, PYGL activation by RIP3 appears to contribute to necrosis.

GLUL is a cytosolic enzyme that catalyzes the condensation of glutamate (Glu) and ammo-
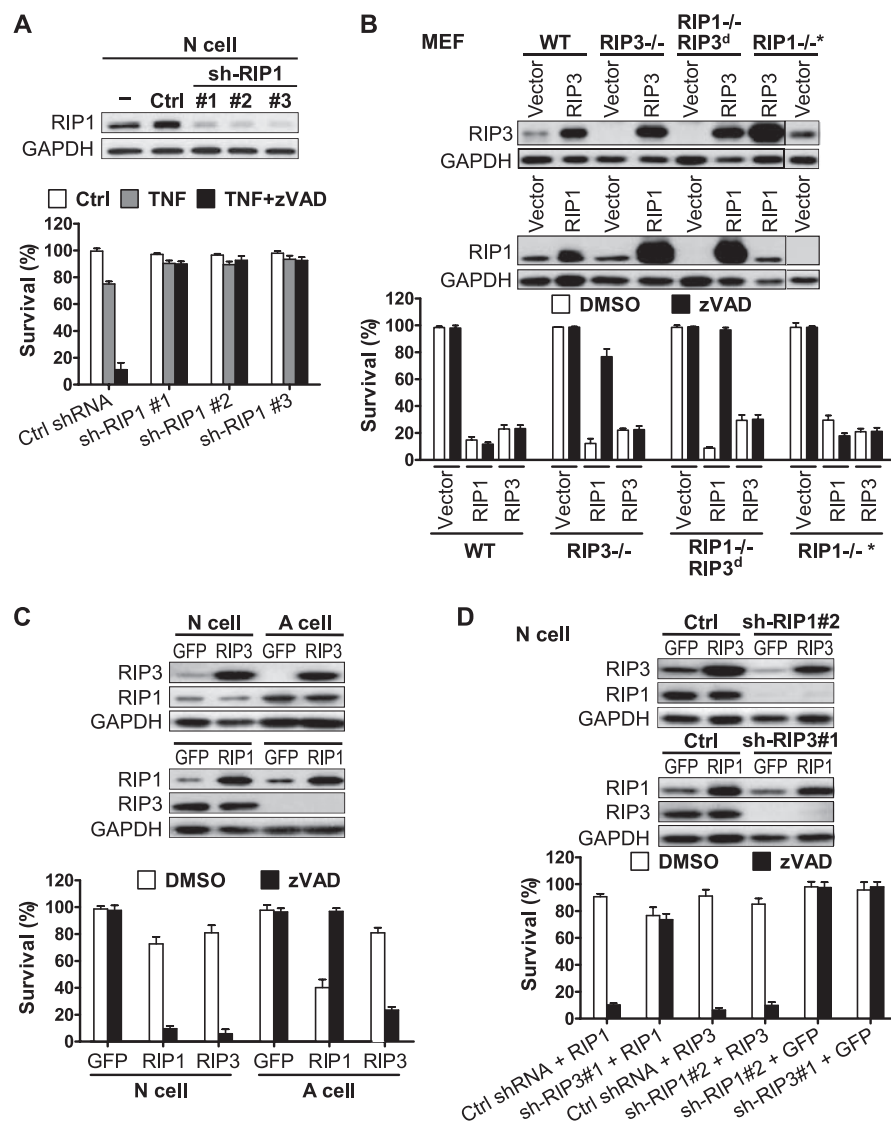


**Fig. 2.** Role of RIP3 and RIP1 in apoptosis and necrosis. (**A**) Viabilities of N cells treated with or without (–) a lentiviral vector encoding control or RIP1-shRNAs for 48 hours then stimulated with or without TNF or TNF+zVAD for 24 hours. RIP1 protein amounts were analyzed by means of immunoblotting 48 hours after infection. (**B**) Wild-type, RIP3$^{-/-}$, RIP1$^{-/-}$*, and RIP1$^{-/-}$RIP3$^{d}$ MEF cells infected with a lentiviral vector encoding RIP1 or RIP3, treated with or without zVAD. Viabilities and RIP1 and RIP3 expression were analyzed 36 hours after infection. (**C**) N and A cells were treated and analyzed as described in (B). (**D**) RIP1 and RIP3 expression levels and viabilities of N cells infected with lentiviral vectors expressing control shRNA and RIP3; RIP1-shRNA and RIP3; control shRNA and RIP1; or RIP3-shRNA and RIP1 and treated with or without zVAD. Data are the mean ± SD of triplicates and are representative of two to three experiments [(A) to (D)].

nia to form glutamine (Gln). Gln can transfer into the mitochondria to function as an energy substrate. GLUD1 is a mitochondrial matrix enzyme that converts Glu to α-ketoglutarate. GLUL and GLUD1 are essential for the use of amino acid Glu or Gln as substrates for adenosine triphosphate production by means of oxidative phosphorylation. Enzymatic pathways specifically used in the mitochondrial oxidation of Gln have been suggested to sensitize the mitochondria to TNF-induced perturbation (20). We detected increased interaction of endogenous RIP3 with GLUL and GLUD1 after N cells were treated with TNF or TNF+zVAD (figs. S7, A and B, and S8, A and B). Recombinant RIP3, but not RIP3-K51A, can directly increase GLUL activity

(fig. S7C). We were unable to prepare a functional recombinant protein of GLUD1, but coexpression of RIP3 with GLUD1 in 293T cells enhanced activity of GLUD1 (fig. S8C). We were unable to detect the activities of endogenous GLUL and GLUD1 in N cells, and therefore were unable to address their activation in cells. Deletion of either GLUL or GLUD1 by means of siRNA reduced TNF+zVAD–induced cell death in N cells (figs. S7D and S8D), suggesting that GLUL- and GLUD1-mediated use of Glu or Gln as energy substrates contributes to necrotic cell death.

Because ROS production is required for TNF- or TNF+zVAD–induced necrosis in L929 cells, MEFs, and macrophages (6, 9, 19), we hypothesized that RIP3 might increase energy

metabolism–associated ROS production. We confirmed that ROS is required for necrosis of N cells but not apoptosis of A cells (fig. S9). We found that depletion of RIP3 in N cells reduced ROS concentration in TNF+zVAD–treated cells (Fig. 4A) and that increasing RIP3 levels in A cells increased TNF+zVAD–induced ROS production (Fig. 4B). RIP3 dependence of ROS production was also confirmed by use of RIP3$^{-/-}$ macrophages (fig. S10). In all cases, the amount of ROS correlated with caspase-independent cell death (Fig. 1, D to F). Depletion of PYGL, GLUL, and GLUD1 by means of siRNA reduced TNF+zVAD–induced accumulation of ROS (Fig. 4C and fig. S11), which correlated with cell death (Fig. 3F and figs. S7D and S8D). Experiments performed by use of energy metabolism inhibitors suggested that mitochondrial ROS generated at the ubisemiquinone site has a key role in TNF cytotoxicity in N cells (fig. S12), which is consistent with a previous report on L929 cells (21). Because increasing glucose by breaking down glycogen and promoting the use of Glu and Gln as energy substrates all function upstream of ubisemiquinone so as to increase energy metabolism, the role of RIP3 in apoptosis/necrosis switching should at least partly occur through increasing energy metabolism–associated ROS production.

A metabolic increase may be needed when cells are treated with a pleiotropic cytokine such as TNF. In cells that have sufficient RIP3 expression, the gateways to using glycogen and Gln or Glu are readily opened after TNF stimulation. zVAD may enhance RIP3 function because RIP3 and RIP1 are reported to be cleaved by caspases (fig. S13) (8, 14). The enhanced metabolism should be accompanied by increased ROS production, which is probably responsible in part for the function of RIP3 in mediating necrosis (Fig. 4D). The role of RIP3 in determining the type of cell death supports the idea that energy metabolism affects cell-death mechanisms (22, 23).

Necrosis occurs under various physiological and pathological conditions, and some might occur in vivo when caspases are inactivated by S-nitrosylation or other means. Because inhibition of necrosis by a RIP1 inhibitor has a beneficial effect on ischemic brain injury and other animal disease models (3, 24), inhibition of RIP3 could be beneficial for diseases that are associated with necrosis, such as diabetes and cerebral ischemia. Indeed, deletion of RIP3 improved the condition of acute pancreatitis in a mouse model (table S3). RIP3 is a potential drug target for necrosis-related diseases.



**Fig. 3.** Activation of PYGL by RIP3 and the involvement of PYGL in necrosis. (**A**) A stable HA-PYGL–expressing N cell line was treated with TNF+zVAD for various time intervals. The cell lysates and antibody-to-RIP3 immunoprecipitates were immunoblotted with antibodies to HA and RIP3. (**B**) Same as in (A), except the treatments were nothing, zVAD, TNF, or TNF+zVAD. (**C**) PYGL activity after flag-PYGL was incubated with 0.25 mM adenosine monophosphate (AMP) for 10 min then with or without recombinant GST-RIP3 or GST-RIP3-K51A (inactive mutant) in a kinase buffer for 30 min at 30°C. (**D**) PYGL activity in N cells infected with a lentiviral vector expressing control shRNA or RIP3 shRNA for 48 hours, treated with nothing, TNF, or TNF+zVAD for 2 hours. (**E**) PYGL activity in wild-type and RIP3$^{-/-}$ mouse peritoneal macrophages, treated with nothing, TNF+zVAD, or LPS+zVAD for 2 hours. (**F**) Viabilities of N cells infected with a lentivirus expressing control shRNA or PYGL shRNAs for 48 hours then stimulated with or without TNF+zVAD for 24 hours. PYGL mRNA levels were determined by means of real-time polymerase chain reaction 48 hours after infection. Data are the mean ± SD of triplicates [(C) to (F)] and are representative of two experiments [(A) to (F)].

**References and Notes**
1. P. Golstein, G. Kroemer, *Trends Biochem. Sci.* **32**, 37 (2007).
2. D. Wallach, T. B. Kang, A. Kovalenko, *Cell Death Differ.* **15**, 1533 (2008).
3. A. Degterev, J. Yuan, *Nat. Rev. Mol. Cell Biol.* **9**, 378 (2008).
4. S. W. Tait, D. R. Green, *Oncology* **27**, 6452 (2008).
5. N. Holler et al., *Nat. Immunol.* **1**, 489 (2000).
6. Y. Lin et al., *J. Biol. Chem.* **279**, 10822 (2004).
7. D. Vercammen et al., *J. Exp. Med.* **187**, 1477 (1998).

**Fig. 4.** Activation of PYGL, GLUL, and GLUD1 by RIP3 contributes to TNF+zVAD–induced ROS production. (**A**) ROS levels in N cells treated with a lentiviral vector expressing control shRNA or RIP3-shRNAs for 48 hours then stimulated with TNF+zVAD for 12 hours. (**B**) ROS levels in A cells infected with or without a lentivirus encoding nothing or RIP3 for 36 hours then stimulated with TNF+zVAD for 6 hours. (**C**) ROS levels in N cells infected with a lentivirus expressing control shRNA, PYGL-shRNA, GLUL-shRNA, or GLUD1-shRNA for 48 hours then stimulated with TNF+zVAD for 12 hours. (**D**) Proposed mechanism for RIP3 in necrosis. When RIP3 is absent, RIP1-mediated cell death is apoptotic. High levels of RIP3 switch apoptosis to necrosis in some cell systems. RIP3 activates PYGL, increasing the availability of energy substrate glucose; RIP3 also activates GLUL and GLUD1, increasing Glu and Gln consumption as energy substrates. These lead to an increase in energy metabolism and subsequent overproduction of the oxidative metabolism product, ROS. ROS is at least in part responsible for RIP3-mediated necrosis.
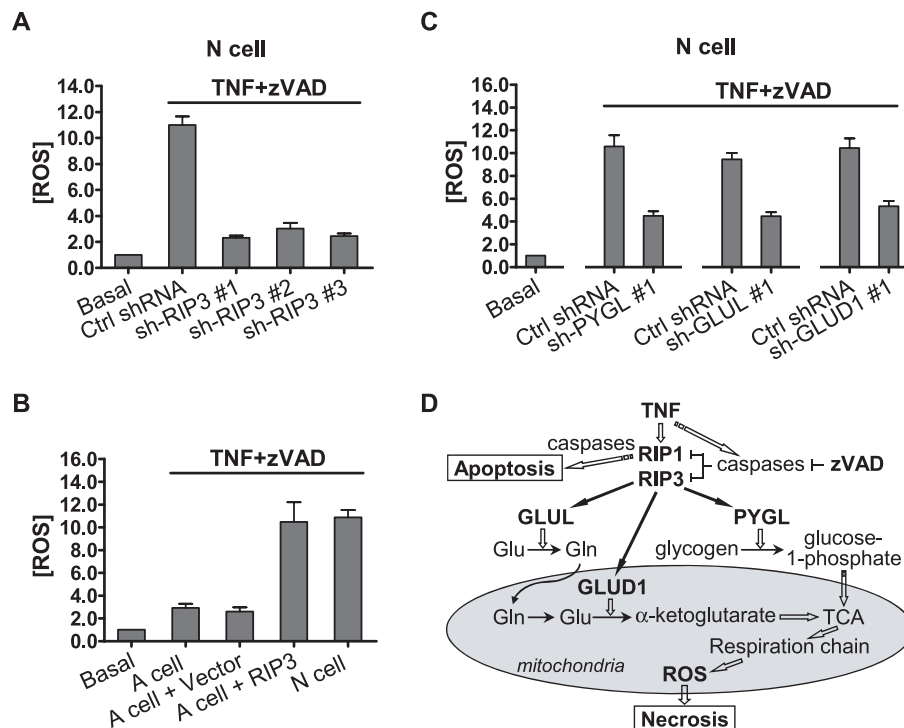
8. F. K. Chan *et al.*, *J. Biol. Chem.* **278**, 51613 (2003).
9. V. Goossens, J. Grooten, K. De Vos, W. Fiers, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8115 (1995).
10. J. Hitomi *et al.*, *Cell* **135**, 1311 (2008).
11. S. Luschen, S. Ussat, G. Scherer, D. Kabelitz, S. Adam-Klages, *J. Biol. Chem.* **275**, 24670 (2000).
12. X. Sun *et al.*, *J. Biol. Chem.* **274**, 16871 (1999).
13. E. Meylan, F. Martinon, M. Thome, M. Gschwendt, J. Tschopp, *EMBO Rep.* **3**, 1201 (2002).
14. S. Feng *et al.*, *Cell. Signal.* **19**, 2056 (2007).
15. G. M. Kasof, J. C. Prosser, D. Liu, M. V. Lorenzi, B. C. Gomes, *FEBS Lett.* **473**, 285 (2000).
16. K. Newton, X. Sun, V. M. Dixit, *Mol. Cell. Biol.* **24**, 1464 (2004).
17. Materials and methods are available as supporting material on *Science* Online.
18. V. Temkin, Q. Huang, H. Liu, H. Osada, R. M. Pope, *Mol. Cell. Biol.* **26**, 2215 (2006).
19. S. O. Kim, K. Ono, P. S. Tobias, J. Han, *J. Exp. Med.* **197**, 1441 (2003).
20. V. Goossens, J. Grooten, W. Fiers, *J. Biol. Chem.* **271**, 192 (1996).
21. K. Schulze-Osthoff *et al.*, *J. Biol. Chem.* **267**, 5317 (1992).
22. N. Festjens, B. T. Vanden, P. Vandenabeele, *Biochim. Biophys. Acta* **1757**, 1371 (2006).
23. W. X. Zong, C. B. Thompson, *Genes Dev.* **20**, 1 (2006).
24. A. Degterev *et al.*, *Nat. Chem. Biol.* **1**, 112 (2005).
25. We thank K. Newton and V. M. Dixit for RIP3$^{-/-}$ mice. This work was supported by grants NSFC-CIHR 30611120526 and 973 program 2009CB522200.

# Reversal of RNA Dominance by Displacement of Protein Sequestered on Triplet Repeat RNA

Thurman M. Wheeler,[1] Krzysztof Sobczak,[1] John D. Lueck,[2] Robert J. Osborne,[1] Xiaoyan Lin,[1] Robert T. Dirksen,[2] Charles A. Thornton[1]*

Genomic expansions of simple tandem repeats can give rise to toxic RNAs that contain expanded repeats. In myotonic dystrophy, the expression of expanded CUG repeats (CUG$^{exp}$) causes abnormal regulation of alternative splicing and neuromuscular dysfunction. We used a transgenic mouse model to show that derangements of myotonic dystrophy are reversed by a morpholino antisense oligonucleotide, CAG25, that binds to CUG$^{exp}$ RNA and blocks its interaction with muscleblind-like 1 (MBNL1), a CUG$^{exp}$-binding protein. CAG25 disperses nuclear foci of CUG$^{exp}$ RNA and reduces the overall burden of this toxic RNA. As MBNL1 is released from sequestration, the defect of alternative splicing regulation is corrected, thereby restoring ion channel function. These findings suggest an alternative use of antisense methods, to inhibit deleterious interactions of proteins with pathogenic RNAs.

Myotonic dystrophy type 1 (DM1) is representative of a group of dominantly inherited disorders in which expression of a toxic RNA leads to neuromuscular degeneration (*1–5*). A feature common to these pathogenic RNAs is the presence of an expanded repeat. In DM1, the disease-inducing transcript is the DM protein kinase (*DMPK*) mRNA containing an expanded CUG repeat in its 3′ untranslated region (3′UTR) (*6*). These CUG repeats bind to muscleblind-like 1 (MBNL1), a splicing regulator, with high affinity (*7, 8*). Because each mutant transcript typically contains thousands of CUG repeats, the capacity for protein binding is large, causing MBNL1 to become sequestered in ribonucleoprotein complexes. These complexes are observed in DM1 cells as foci of CUG$^{exp}$-MBNL1 aggregates in the nucleus (*7, 9–11*).

If sequestration of MBNL1 contributes to symptoms of DM1, inhibitors of MBNL1-CUG$^{exp}$ binding may reverse these effects. To test this possibility, we used CAG25, an antisense 25-nucleotide morpholino oligonucleotide composed of CAG repeats. Antisense morpholinos do not trigger cleavage of their target RNAs (*12*), which suggests that they could be used to bind CUG$^{exp}$ RNA and release sequestered proteins, without risk of degrading other transcripts that contain CUG repeats. A potential obstacle for this approach, however, is that CUG$^{exp}$ RNAs form

hairpins with high thermal stability (13, 14) and extensive MBNL1 binding (8), which potentially could limit their access to antisense oligonucleotides. In vitro CAG25 was able to invade CUG$^{exp}$ hairpins and form a stable RNA-morpholino heteroduplex (Fig. 1A and figs. S1 and S2). CAG25 was also able to block the formation of CUG$^{exp}$-MBNL1 complexes and disrupt complexes that had already formed (Fig. 1, B to D).

To examine whether CAG25 can influence CUG$^{exp}$ interactions in vivo, we tested its effects in a transgenic mouse model of DM1. $HSA^{LR}$ transgenic mice express human skeletal actin transcripts that have $(CUG)_{250}$ inserted in the 3'UTR. These mice accumulate CUG$^{exp}$ RNA

[1]Departments of Neurology, Pharmacology and Physiology, University of Rochester, Rochester, NY 14642, USA. [2]Departments of Pharmacology and Physiology, University of Rochester, Rochester, NY 14642, USA.

*To whom correspondence should be addressed. E-mail: charles_thornton@urmc.rochester.edu

and MBNL1 protein in nuclear foci in skeletal muscle (1), a process that is thought to depend on CUG$^{exp}$-MBNL1 interaction (15). We therefore examined the effect of CAG25 on foci in muscle cells. We loaded CAG25 into muscle fibers by intramuscular injection followed by in vivo electroporation. Muscle tissue was examined 1 to 3 weeks later by means of fluorescence in situ hybridization, using probes that hybridize to the CUG repeat or to sequences flanking the repeat. Injection of CAG25, but not a control morpholino of unrelated sequence, caused a marked reduction of nuclear foci and a redistribution of MBNL1 protein (Fig. 1, E and F, and fig. S3).

To determine whether CAG25 can reverse the biochemical consequences of MBNL1 sequestration, we examined its effects on alternative splicing. $HSA^{LR}$ transgenic mice show alternative splicing changes similar to those observed in human DM1 (10). The splicing misregulation is

improved when MBNL1 levels are increased (16), aggravated when MBNL1 levels are reduced, and reproduced by ablation of $Mbnl1$ (17), suggesting that splicing defects in this model are primarily caused by MBNL1 sequestration. For each DM1-affected exon that we examined, the alternative splicing was normalized or nearly corrected at 3 weeks after injection of CAG25 (Fig. 2). Effects of CAG25 on alternative splicing persisted at 14 weeks (fig. S4, A and B) but not at 8 months after a single injection. In contrast, CAG25 did not correct the misregulated splicing of these same exons in $Mbnl1$ knockout mice or alter their splicing patterns in wild-type (WT) mice (fig. S5), indicating that its effects are mediated through CUG$^{exp}$ RNA rather than acting directly on the respective precursor mRNAs (pre-mRNAs). The $Capzb$ and $Itgb1$ transcripts also show developmentally regulated alternative splicing in skeletal muscle, but they do not depend on MBNL1, they are not misregulated in $HSA^{LR}$
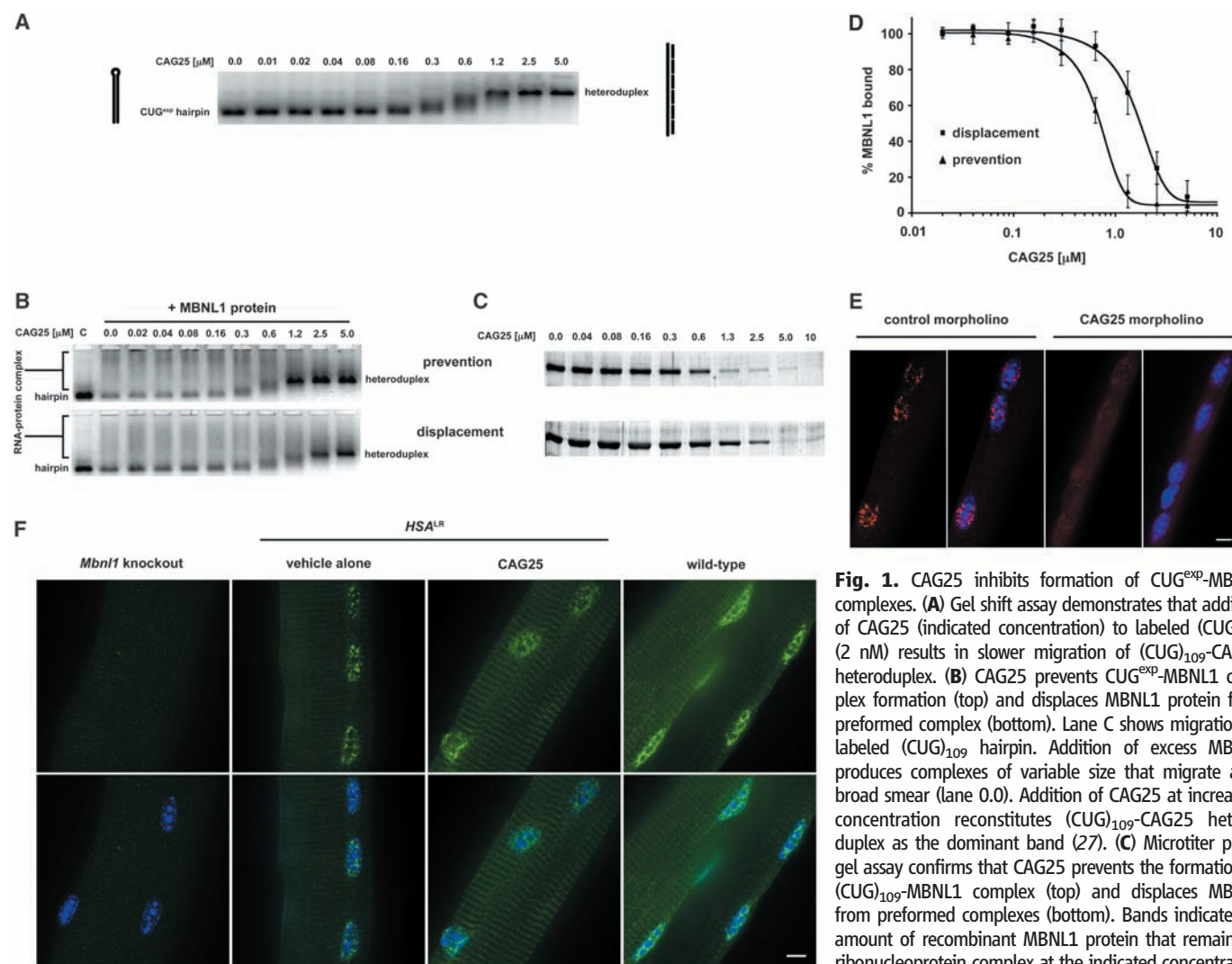


**Fig. 1.** CAG25 inhibits formation of CUG$^{exp}$-MBNL1 complexes. (**A**) Gel shift assay demonstrates that addition of CAG25 (indicated concentration) to labeled $(CUG)_{109}$ (2 nM) results in slower migration of $(CUG)_{109}$-CAG25 heteroduplex. (**B**) CAG25 prevents CUG$^{exp}$-MBNL1 complex formation (top) and displaces MBNL1 protein from preformed complex (bottom). Lane C shows migration of labeled $(CUG)_{109}$ hairpin. Addition of excess MBNL1 produces complexes of variable size that migrate as a broad smear (lane 0.0). Addition of CAG25 at increasing concentration reconstitutes $(CUG)_{109}$-CAG25 heteroduplex as the dominant band (27). (**C**) Microtiter plate/gel assay confirms that CAG25 prevents the formation of $(CUG)_{109}$-MBNL1 complex (top) and displaces MBNL1 from preformed complexes (bottom). Bands indicate the amount of recombinant MBNL1 protein that remains in ribonucleoprotein complex at the indicated concentration of CAG25 (27). (**D**) The percent of MBNL1 bound to CUG$^{exp}$ is expressed as the mean ± SD of protein retained on plate. Median inhibitory concentration (IC$_{50}$) for "prevention" is 462 ± 31 nM and for "displacement" is 1032 ± 117 nM. (**E**) Fluorescence in situ hybridization of single flexor digitorum brevis (FDB) muscle fibers from $HSA^{LR}$ mice. Probe (red) binds to $HSA^{LR}$ transcripts upstream from CUG repeat; nuclei are blue. CAG25, but not control morpholino, causes dispersal of RNA foci. (**F**) MBNL1 (immunofluorescence, green) shifts from punctate to diffuse nuclear distribution after injection of FDB with CAG25. Scale bars, 5 μm.
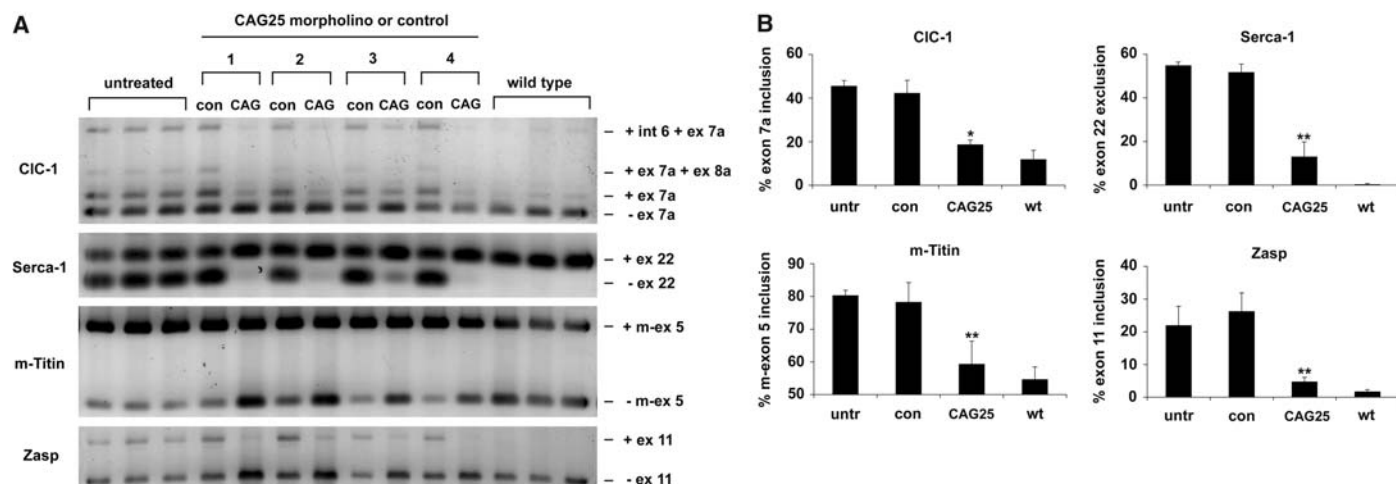
**Fig. 2.** Reversal of misregulated alternative splicing by CAG25. (**A**) Reverse transcription polymerase chain reaction (RT-PCR) analysis of alternative splicing for ClC-1, Serca-1, m-Titin, and Zasp, 3 weeks after a single injection of CAG25 in tibialis anterior (TA) muscle of $HSA^{LR}$ transgenic mice. The contralateral (con) TA was injected with vehicle (saline, mice 1 and 2) or morpholino with inverted sequence (GAC25, mice 3 and 4). Splice products from untreated $HSA^{LR}$ transgenic and WT TA muscle ($n = 3$ mice) are shown. Int 6, intron 6 retention. (**B**) Quantification of results in (A), expressed as the percentage of splice products that include or exclude the indicated exon. *$P = 0.003$ and **$P < 0.001$ for CAG25 versus contralateral (Student's $t$ test). untr, untreated $HSA^{LR}$.

**Fig. 3.** ClC-1 protein expression and function are rescued by CAG25. (**A**) Immunofluorescence for ClC-1 protein expression in sections from $HSA^{LR}$ TA muscle. Scale bar, 20 μm. (**B**) ClC-1 current density in FDB fibers isolated from 15-day-old $HSA^{LR}$ mice injected with CAG25 or control morpholino. Fibers from WT mice injected with CAG25 or control morpholino serve as controls (fig. S6). (**C**) Electromyographic myotonia analysis 3 weeks after injection of CAG25 into TA muscle. The contralateral side was injected with vehicle (saline) alone or control morpholino (27). $n = 11$ mice examined; *$P < 0.0001$ for CAG25 versus saline (Student's $t$ test).
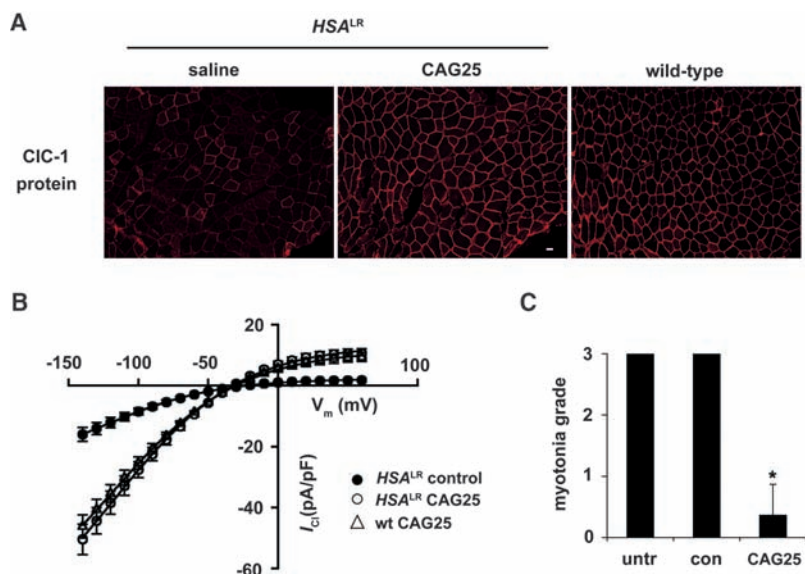


mice (10), and CAG25 had no effect on splicing of either transcript.

To determine whether CAG25 can rescue the physiological deficits of DM1, we examined the expression and function of chloride channel 1 (ClC-1), the muscle-specific chloride channel. Delayed muscle relaxation and repetitive action potentials (myotonia) are cardinal features of DM1, resulting from loss of ClC-1 channels (18, 19). Affected individuals and mouse models show inclusion of an additional exon in the ClC-1 mRNA, resulting in frameshift and loss of channel activity (20). CAG25 corrected the defect of ClC-1 alternative splicing (Fig. 2 and fig. S4, A and B) and restored the expression of ClC-1 protein to the surface membrane (Fig. 3A and fig. S4C). Furthermore, transmembrane chloride ion conductance was normalized (Fig. 3B), and myotonia

was markedly reduced (Fig. 3C and figs. S4D and S7).

The effects of MBNL1 sequestration on gene expression are not limited to alternative splicing. For example, transcription of *Eda2r*, *Uchl1*, and *Sarcolipin* is highly upregulated in *Mbnl1* knockout mice, and similar changes are induced by the expression of CUG$^{exp}$ (21). Although the mechanism for this effect has not been determined, it was partially reversed by CAG25 (fig. S6). These data indicate that CAG25 ameliorates both transcriptional and posttranscriptional effects of toxic RNA.

We next determined whether CAG25 can overcome the nuclear retention of CUG-expanded transcripts (22). Despite reducing the overall level of CUG$^{exp}$ RNA in muscle (see below), CAG25 increased the amount of this transcript in the cyto-

plasm (Fig. 4A and fig. S8). Moreover, because reduced translation of DMPK mRNA may contribute to cardiac symptoms of DM1 (23), we also tested whether CAG25 could enhance the translation of CUG$^{exp}$-containing transcripts. We derived transgenic mice that express luciferase mRNA containing $(CUG)_{270}$ in the 3'UTR. The luciferase transcript is retained in nuclear foci, and basal levels of luciferase activity are accordingly reduced. In vivo bioluminescence imaging showed that CAG25 induced a focal increase of luciferase activity in the injected hindlimb (Fig. 4B), which is consistent with increased translation of the CUG-expanded mRNA.

CAG25 caused a ~50% reduction in the overall burden of CUG$^{exp}$ RNA (Fig. 4C). There was no parallel reduction of transgene pre-mRNA (fig. S9), suggesting that downregulation of this
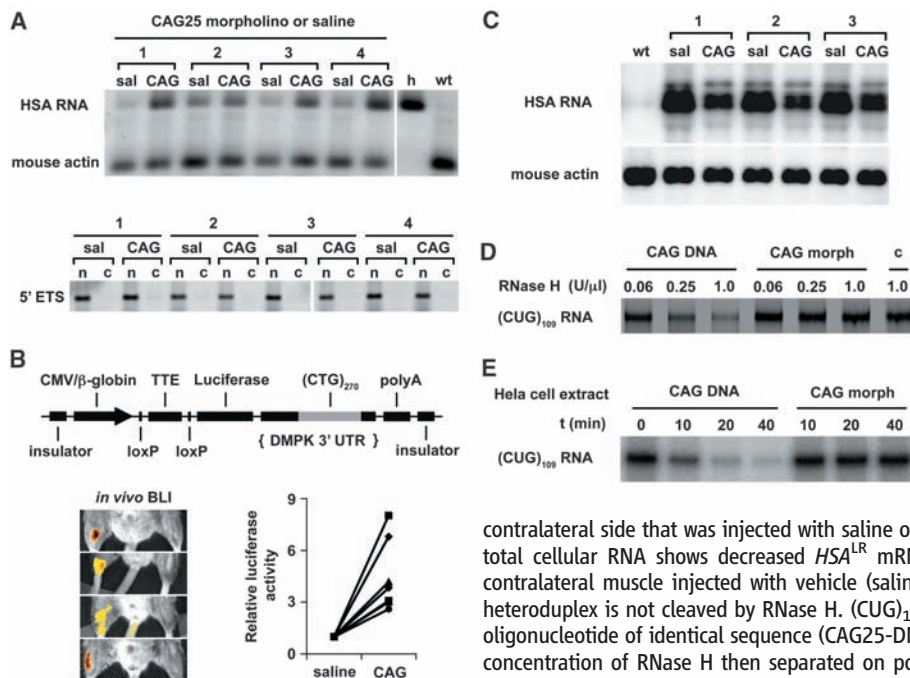
**Fig. 4.** CAG25 increases the cytoplasmic level and translation of CUG$^{exp}$-containing mRNA despite reducing the overall level of CUG$^{exp}$ RNA. (**A**) RT-PCR assay of transgene mRNA in the cytoplasm. Human (transgene) and mouse (endogenous) skeletal actin transcripts were coamplified by the same primers; the species origin was revealed with AluI cleavage (27). h, human muscle. (Bottom) Depletion of nuclear RNA from cytoplasmic fraction (c) relative to nuclear pellet (n), analyzed by means of RT-PCR for the 5' external transcribed spacer (5'ETS) (nuclear-retained) of ribosomal RNA (rRNA). Numbers refer to different mice treated with CAG25. (**B**) CAG25 increases luciferase activity in LLC9/Rosa-CreER bitransgenic mice. (Top) LLC9 transgene for conditional expression of CUG$^{exp}$ RNA. (Bottom) In vivo bioluminescence imaging (BLI) of different bitransgenic mice (27). For quantification, luciferase activity (indicated in yellow and orange) in CAG25-injected muscle was normalized to the contralateral side that was injected with saline or control morpholino (n = 7 mice). (**C**) Northern blot of total cellular RNA shows decreased HSA$^{LR}$ mRNA in muscle injected with CAG25 as compared with contralateral muscle injected with vehicle (saline). Mouse actin is loading control. (**D**) CUG$^{exp}$-CAG25 heteroduplex is not cleaved by RNase H. (CUG)$_{109}$ RNA was incubated with CAG25 morpholino or DNA oligonucleotide of identical sequence (CAG25-DNA). The heteroduplex was incubated with the indicated concentration of RNase H then separated on polyacrylamide gels. Lane c is the control with (CUG)$_{109}$ alone. (**E**) (CUG)$_{109}$ RNA was incubated with CAG25 morpholino or CAG25-DNA to form heteroduplex as in (D), to which HeLa cell extract was added for the indicated time.

transcript results from accelerated decay. However, previous work has shown that antisense morpholinos do not support RNA cleavage by ribonuclease H (RNase H) (12). Consistent with these observations, CAG25 did not induce cleavage of CUG$^{exp}$ RNA by recombinant RNase H or by HeLa cell extracts in vitro, whereas an equivalent DNA oligonucleotide caused extensive CUG$^{exp}$ degradation (Fig. 4, D and E). Moreover, CAG25 did not reduce levels of endogenous CUG-repeat-containing transcripts (fig. S10), nor did it reduce transgene mRNA in mice that express an equivalent HSA transcript containing a nonexpanded CUG repeat (fig. S11). Taken together, these results are consistent with an indirect effect of CAG25 on the accumulation of CUG$^{exp}$ RNA. However, therapeutic effects of CAG25 do not derive solely from downregulation of the repetitive RNA because the residual levels of CUG$^{exp}$ were not below the threshold for inducing RNA disease (fig. S12).

DM1 presents a complex phenotype that results from trans-dominant effects of mutant RNA on many different transcripts. To intervene at an upstream site and accomplish a general correction of DM1, we used an antisense methodology to inhibit the protein interactions of expanded RNA repeats. This strategy has several effects that are potentially beneficial, including the release of MBNL1 protein from ribonucleoprotein foci, enhanced transport of CUG$^{exp}$ transcripts to the cytoplasm, and a reduced burden of CUG$^{exp}$ RNA. The mechanism for the latter effect remains to be determined, but we postulate that it may result from accelerated decay of CUG$^{exp}$ RNA once it has been transported to the cytoplasm. Considering that the extent of MBNL1 sequestration in DM1 is variable in different nuclei from the

same individual (10), and that ~50% of normal MBNL1 levels are sufficient to ensure normal splicing regulation in mice (10, 17), even partial release of sequestered MBNL1 probably improves splicing abnormalities in DM1. The mass of CUG$^{exp}$ RNA in muscle from HSA$^{LR}$ mice is two- to eightfold higher than in muscle tissue from patients with DM1 (fig. S13), suggesting that similar therapeutic effects can be achieved in humans if the antisense can be effectively delivered. Indeed, treatment with CAG25 of DM1 cells in tissue culture led to a reduction of intranuclear RNA foci (fig. S14). Among several CUG$^{exp}$-protein interactions that may contribute to DM1 pathogenesis, we have focused on MBNL1 because it has the highest CUG$^{exp}$-binding affinity (8, 24) and most complete sequestration (11) of any factor so far identified. However, it seems likely that interactions of CUG$^{exp}$ with other RNA-binding proteins, such as CUG-binding protein 1 (25), will also be inhibited by this approach, which may favorably affect signaling abnormalities in DM1 (26). Taken together, these data supply proof of concept that agents inhibiting deleterious RNA-protein interactions have therapeutic potential in RNA-dominant disorders.

### References and Notes

1. A. Mankodi et al., Science **289**, 1769 (2000).
2. C. L. Liquori et al., Science **293**, 864 (2001).
3. P. Jin et al., Neuron **39**, 739 (2003).
4. M. L. Moseley et al., Nat. Genet. **38**, 758 (2006).
5. L. B. Li, Z. Yu, X. Teng, N. M. Bonini, Nature **453**, 1107 (2008).
6. J. D. Brook et al., Cell **68**, 799 (1992).
7. J. W. Miller et al., EMBO J. **19**, 4439 (2000).
8. Y. Yuan et al., Nucleic Acids Res. **35**, 5474 (2007).
9. K. L. Taneja, M. McCurrach, M. Schalling, D. Housman, R. H. Singer, J. Cell Biol. **128**, 995 (1995).
10. X. Lin et al., Hum. Mol. Genet. **15**, 2087 (2006).
11. H. Jiang, A. Mankodi, M. S. Swanson, R. T. Moxley, C. A. Thornton, Hum. Mol. Genet. **13**, 3079 (2004).
12. J. Summerton, Biochim. Biophys. Acta **1489**, 141 (1999).
13. M. Napierala, W. J. Krzyzosiak, J. Biol. Chem. **272**, 31079 (1997).
14. B. Tian et al., RNA **6**, 79 (2000).
15. W. Dansithong, S. Paul, L. Comai, S. Reddy, J. Biol. Chem. **280**, 5773 (2005).
16. R. N. Kanadia et al., Proc. Natl. Acad. Sci. U.S.A. **103**, 11748 (2006).
17. R. N. Kanadia et al., Science **302**, 1978 (2003).
18. A. Mankodi et al., Mol. Cell **10**, 35 (2002).
19. N. Charlet-B. et al., Mol. Cell **10**, 45 (2002).
20. J. D. Lueck, A. Mankodi, M. S. Swanson, C. A. Thornton, R. T. Dirksen, J. Gen. Physiol. **129**, 79 (2007).
21. R. J. Osborne et al., Hum. Mol. Genet. **18**, 1471 (2009).
22. B. M. Davis, M. E. McCurrach, K. L. Taneja, R. H. Singer, D. E. Housman, Proc. Natl. Acad. Sci. U.S.A. **94**, 7388 (1997).
23. C. I. Berul et al., J. Clin. Invest. **103**, R1 (1999).
24. M. B. Warf, J. A. Berglund, RNA **13**, 2238 (2007).
25. L. T. Timchenko et al., Nucleic Acids Res. **24**, 4407 (1996).
26. N. M. Kuyumcu-Martinez, G. S. Wang, T. A. Cooper, Mol. Cell **28**, 68 (2007).
27. Materials and methods are available as supporting material on Science Online.
28. This work comes from the University of Rochester Wellstone Muscular Dystrophy Cooperative Research Center (U54NS48843) and Center for RNA Biology, with support from NIH (AR046806, AR/NS48143, and NIDCR-T32DE07202 to J.D.L.), the Muscular Dystrophy Association, and a postdoctoral fellowship to K.S. from the Foundation for Polish Science, Run America Foundation, and the Saunders Family fund. We appreciate help from L. Richardson and S. Leistman. The University of Rochester has applied for a patent based partly on the work described here.

# Genome-Wide RNAi Screen Identifies Genes Involved in Intestinal Pathogenic Bacterial Infection

Shane J. F. Cronin,[1]* Nadine T. Nehme,[2]* Stefanie Limmer,[2] Samuel Liegeois,[2] J. Andrew Pospisilik,[1] Daniel Schramek,[1] Andreas Leibbrandt,[1] Ricardo de Matos Simoes,[3] Susanne Gruber,[1] Urszula Puc,[1] Ingo Ebersberger,[3] Tamara Zoranovic,[1] G. Gregory Neely,[1] Arndt von Haeseler,[3] Dominique Ferrandon,[2]†‡ Josef M. Penninger[1]†

Innate immunity represents the first line of defense in animals. We report a genome-wide in vivo *Drosophila* RNA interference screen to uncover genes involved in susceptibility or resistance to intestinal infection with the bacterium *Serratia marcescens*. We first employed whole-organism gene suppression, followed by tissue-specific silencing in gut epithelium or hemocytes to identify several hundred genes involved in intestinal antibacterial immunity. Among the pathways identified, we showed that the JAK-STAT signaling pathway controls host defense in the gut by regulating stem cell proliferation and thus epithelial cell homeostasis. Therefore, we revealed multiple genes involved in antibacterial defense and the regulation of innate immunity.

*D*rosophila melanogaster provides a powerful model that allows the dissection of the innate immune response at the organism level. In *Drosophila*, innate immunity has a humoral and a cellular immune response. The majority of our knowledge of *Drosophila* immunity is based on injection of nonpathogenic bacteria (*1–3*); however, this bypasses the initial steps of naturally occurring infections—namely, the physical barriers and the local, mucosal immune response. Intestinal immunity is currently the focus of intense research (*4*). In contrast to the human digestive tract, *Drosophila* lacks mammalian-like adaptive immunity and so relies entirely upon an innate immune system for protection against invading pathogens.

The intestinal infection model using pathogenic *Serratia marcescens* allows for the detailed analysis of local intestinal immunity and phagocytosis (*5*). *S. marcescens* is a gram-negative, opportunistic pathogen that can infect a range of hosts including *Drosophila, Caenorhabditis elegans,* and mammals (*6, 7*). Using ubiquitous RNA interference (RNAi)–mediated suppression, we performed an inducible genome-wide in vivo screen in *Drosophila* for novel innate immune regulators after *S. marcescens* infection (*8*) [Fig. 1A, fig. S1, A and B, and supporting online material (SOM) text]. To confirm our experimental approach we assayed various members of the Immune deficiency (IMD) and Toll pathways, the two major fly immune signaling cascades (Fig. 1B) (*1–3*). RNAi lines targeting several

IMD members resulted in significantly reduced survival on infection with *S. marcescens,* whereas suppression of Toll pathway components had a less dramatic effect, which supports previous reports that the immune response to *S. marcescens* is IMD-dependent and Toll-independent (Fig. 1B) (*5*). Notably, not all members of the IMD pathway, such as *imd*, *rel*, and *ird5*, were picked up by our screening criteria, most likely because of inefficient RNAi silencing (Fig. 1B) (*9*).

We assayed 13,053 RNAi lines (*9*) representing 10,689 different genes (78% of the genome) against intestinal infection with *S. marcescens* (fig. S2A and tables S1 and S2). Of these, 8.3% (885 genes) were defined as hits, the majority of which (89.3%; 790 genes) were susceptible candidates (fig. S2A and table S3). On the basis of gene ontology (GO) annotations, susceptible candidates were classified according to their predicted biological processes. Genes involved in signaling, intracellular protein transport, and transcriptional regulation were overly represented among the entire data set (Fig. 1C). We also found marked enrichment for genes that regulate phagocytosis, defense responses, vesicle trafficking, and proteolysis. Several candidate RNAi lines represented genes that have been previously implicated in mounting an effective immune response (*10–19*) (table S3).

Our approach also allowed us to identify negative regulators of *Drosophila* host defense (Fig. 1A). We identified 95 genes (10.7% of the total hits) that confer resistance to *S. marcescens* infections when silenced (fig. S2, A and B, table S4), none of which had previously been characterized as negative regulators of innate immunity. Thus, our genome-wide screen revealed previously

---

[1]IMBA, Institute of Molecular Biotechnology of the Austrian Academy of Sciences, A-1030 Vienna, Austria. [2]Equipe Fondation Recherche Médicale, UPR 9022 du CNRS, Institut de Biologie Moléculaire et Cellulaire du CNRS, F-67084 Strasbourg, France. [3]Center for Integrative Bioinformatics (CIBIV), University of Vienna, Medical University of Vienna, University of Veterinary Medicine, A-1030 Vienna, Austria.

*These authors contributed equally to this work.
†This work is based on equal contributions from the laboratories of the last two authors.
‡To whom correspondence should be addressed. E-mail: d.ferrandon@ibmc.u-strasbg.fr

**Fig. 1.** Analysis of genome-wide in vivo RNAi screen. (**A**) Total data of all RNAi lines screened for survival after *S. marcescens* infections. Data were analyzed as the time, in days, when 50% of the total number of flies had died. All data were normalized to the daily median time-to-death ($LT_{50}$) mean of an experimental cohort. In all experiments, the cohort ranged from 80 to 200 lines. Hits were defined by susceptible (red dashed line) and resistant (blue dashed line) cut-offs, i.e., 1.5 SD below the mean and 2 SD above the mean, respectively, based on the pilot screen and controls. (**B**) Effect of RNAi knockdown of IMD and Toll pathway components on their survival against *S. marcescens* infection. SCOREs are shown for each line as described in (*8*). The dashed lines indicate the cut-offs used for resistance (+2 SD) and susceptibility (−1.5 SD) candidates. (**C**) Percentage distribution of GO annotated genes to biological processes for susceptible candidates.

known genes associated with *Drosophila* immunity and more than 800 additional candidate genes implicated in innate immunity, 40% of which had unknown function.

We retested some of our susceptible and resistant RNAi hits in the gut epithelium and the macrophage-like hemocytes, the two major cell types associated with our infection model, using cell type–specific driver lines, NP1-GAL4 and HML-GAL4, respectively (*5*, *20*). We prioritized genes of interest by selecting the primary hits that have mammalian (mouse and/or human) orthologs. Of the 358 susceptible hits tested with the HML-GAL4 driver, RNAi against 98 genes (27%) resulted in significantly reduced survival as compared with RNAi controls, which indicated that these genes function in hemocytes to combat intestinal *S. marcescens* infections (Fig. 2A, fig. S3A,

and table S5). When we used the NP1-GAL4 driver (fig. S4) to test 337 genes, RNAi against 129 genes (38%) resulted in significantly reduced survival, which suggested that these genes play an important role in host intestinal defense (Fig. 2B, fig. S3B, and table S6). Of the resistance hits, 37 HML-GAL4 RNAi candidates (79%) and 28 NP1-GAL4 RNAi candidates (61%) exhibited markedly enhanced survival (Fig. 2, C and D, fig. S3, and tables S7 to S9). Of the candidate genes, 79 functioned in both hemocytes and gut (fig. S3). Multiple susceptibility and resistance genes were tested 3 to 15 independent times, using ≥2 RNAi transformants to exclude position effects and second independent RNAi hairpins to confirm the target gene when available (Fig. 2, A to D, fig. S3, and tables S5 to S8). To exclude a potential developmental phenotype, we have tested most candidate lines by feeding flies on a sugar diet in the absence of bacteria (table S9). Thus, we have identified multiple regulators in hemocytes and/or gut epithelium that confer susceptibility or resistance to *S. marcescens* infections.

Using GO enrichment analysis, we classified our tissue-specific candidates into statistically significant biological processes. In the intestinal tract, intracellular processes such as endocytosis and exocytosis, proteolysis, vesicle-mediated transport, and stress response all appeared significantly enriched (Fig. 2E, figs. S5 to S7, and table S10). We also observed a marked enhancement of genes associated with immune system development, growth, stem cell division, and cell death, which suggested an important role for these processes in the gut during *S. marcescens* infection. In hemocytes, ontology enrichment analysis revealed a strong enrichment in several processes linked to phagocytosis including endocytosis, response to external stimuli, and vesicle trafficking (figs. S8 to S10 and table S11). In both cell types, deregulation of the stress response, as well as amine and/or nitrogen metabolism, resulted in enhanced resistance to *S. marcescens* challenge (Fig. 2E and fig. S8).

We next performed Kegg pathway analysis to identify enriched gene sets that might be involved in *S. marcescens* infections. Kegg profiling on the susceptible genome-wide candidates (table S12) showed the importance of the IMD pathway in our infection model and also pointed to a possible role of Notch and transforming growth factor-β signaling pathways, which have previously been difficult to study in an infection setting because of a lack of adult viable mutants (*21*, *22*). Moreover, our analysis revealed prominent involvement of the Janus kinase–signal transducers and activators of transcription (JAK-STAT) pathway during *S. marcescens* infection. In *Drosophila*, the JAK-STAT pathway plays an important role in hematopoiesis, stress responses, stem cell proliferation, and antiviral immunity, but its role in the defense against natural bacterial pathogens is unknown (*23*–*26*). We therefore sought to validate our analysis and focused on how JAK-STAT signaling regulates the host response during *S. marcescens* infection.
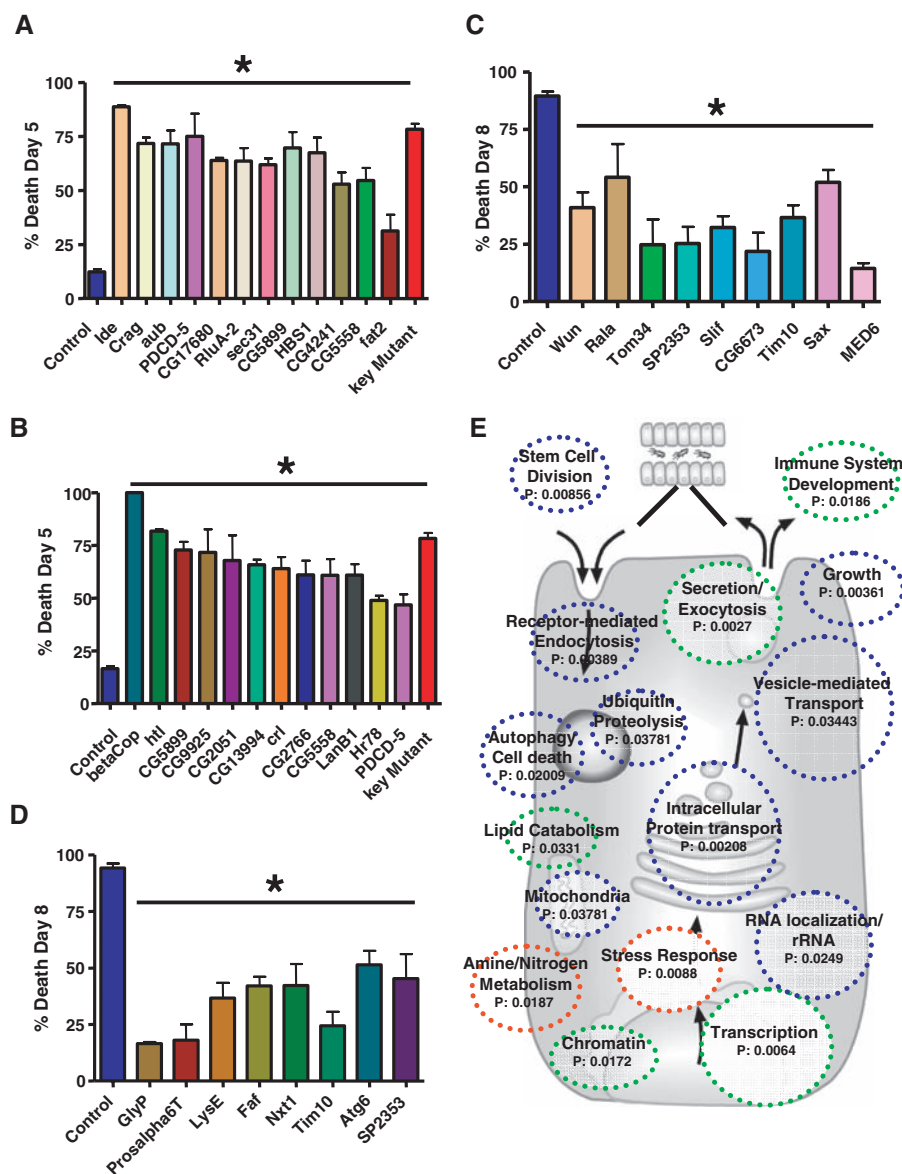


**Fig. 2.** Mapping and validation of conserved hits in the gut and hemocytes. (**A** and **B**) Survival graphs showing susceptible hits, (A) HML-susceptible genes and (B) NP1-susceptible genes, tested 3 to 15 times with several transformants and hairpins in hemocytes and gut epithelium, respectively. The *kenny* mutant line (key Mutant) is shown as a positive control. Means ± SEM, $n ≥ 3$ experiments with 20 flies in each. *$P < 0.05$ (Welch *t* test). (**C** and **D**) Survival graphs showing resistant hits, (C) HML-resistant genes and (D) NP1-resistant genes, tested 3 to 15 times with several transformants and hairpins in hemocytes and gut epithelium, respectively. Means ± SEM, $n ≥ 3$ experiments with 20 flies in each. *$P < 0.05$ (Welch *t* test). (**E**) Statistically enriched biological processes superimposed on a sketch depicting a gut epithelial cell, with the corresponding $P$ value in the gut associated with *S. marcescens* infection. Green indicates processes to which susceptible candidates are exclusively attributed. Red indicates processes to which resistant candidates are exclusively attributed. Blue indicates processes to which both susceptible and resistant candidates can be attributed. See also table S10 for annotation of genes involved in each process. All processes shown display $P < 0.05$ (Fisher's exact test).

To investigate whether the JAK-STAT pathway is activated during *S. marcescens* infection, we used transgenic reporter lines (*24, 27, 28*) in which green fluorescent protein (GFP) is expressed under the control of *unpaired* (*upd*) and *upd-3*, which encode two ligands for Domeless (the receptor of the JAK-STAT pathway). We observed *upd*-GFP and *upd3*-GFP expression in the gut of *S. marcescens*–

infected flies (Fig. 3A and figs. S11 and S12). Moreover, we demonstrated intestinal activation of the JAK-STAT pathway by using a *stat92E*–binding site–GFP reporter line (Fig. 3B) (*27, 28*). On ligation of UPD or UPD3 to Domeless, Stat92E translocates to the nucleus and activates reporter GFP gene expression (*27*). To confirm the relevance of JAK-STAT activation for *S. marcescens* infec-

tions, we performed global (Fig. 3, C and D) and gut-specific (Fig. 3E) RNAi-mediated silencing of PIAS [also called Su(var)-10] and PP1α96A, two negative regulators of JAK-STAT signaling (*29, 30*). In both RNAi lines, we observed significantly earlier death compared with that of control flies (Fig. 3, C to E). The role of PP1α96A in intestinal immunity was also validated using a
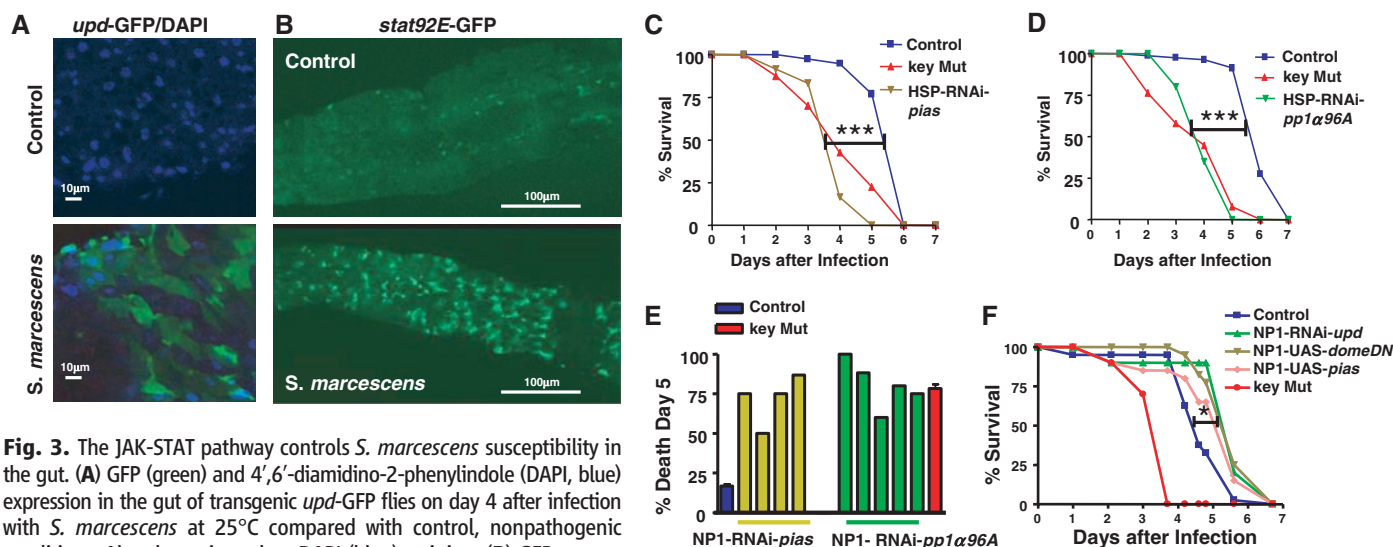


**Fig. 3.** The JAK-STAT pathway controls *S. marcescens* susceptibility in the gut. (**A**) GFP (green) and 4′,6′-diamidino-2-phenylindole (DAPI, blue) expression in the gut of transgenic *upd*-GFP flies on day 4 after infection with *S. marcescens* at 25°C compared with control, nonpathogenic conditions. Also shown is nuclear DAPI (blue) staining. (**B**) GFP expression in the gut of transgenic *stat92E*-GFP flies under *S. marcescens*–infected and control conditions on day 4 at 25°C. (**C**) Survival curves of *S. marcescens*–infected RNAi lines against the negative JAK-STAT pathway regulator PIAS driven by the ubiquitously expressed HSP-GAL4 driver compared with control and *key* mutant flies. (**D**) Survival curves of *S. marcescens*–infected RNAi lines targeting the negative JAK-STAT regulator PP1α96A driven by the ubiquitously expressed HSP-

GAL4 driver compared with control and *key* mutant flies. (**E**) Survival graph representing individual tests of RNAi-mediated silencing of PIAS and PP1α96A specifically in the gut (NP1 driver) after *S. marcescens* challenge at 29°C, compared with control and *key* mutant flies. (**F**) Survival curves of lines shown at 29°C compared with control and *key* mutant flies following *S. marcescens* feeding. *\*P ≤ 0.05; \*\*\*P ≤ 0.0001* (logrank test). *upd*, *unpaired*; DN, dominant-negative.
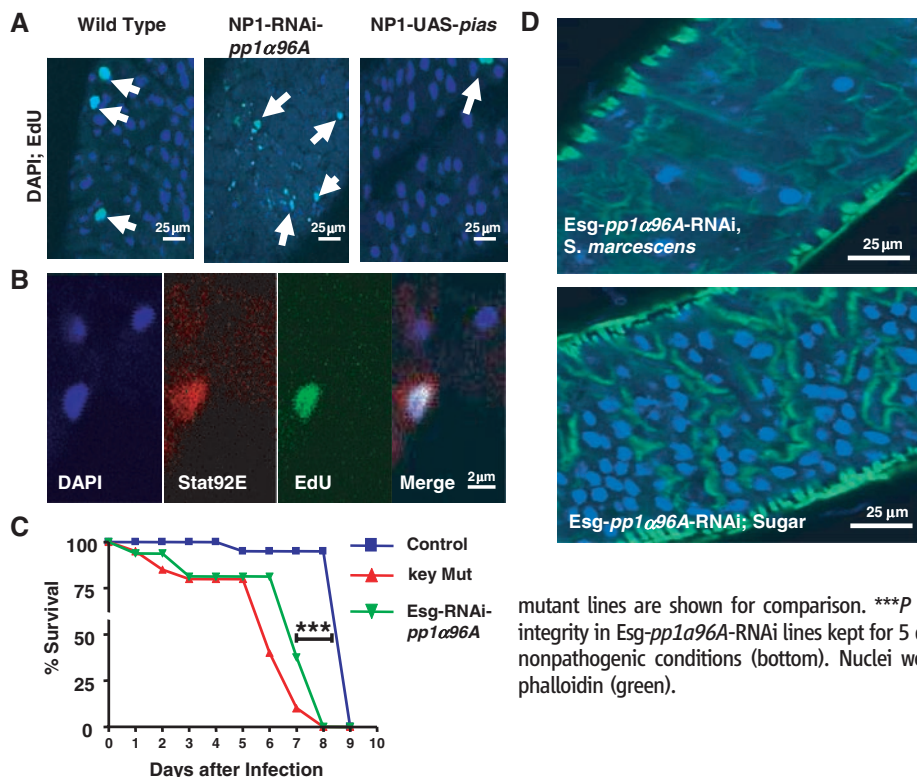


**Fig. 4.** Impaired epithelial integrity and control of intestinal stem cell homeostasis upon *S. marcescens* challenge. (**A**) Analysis of gut epithelium integrity using DAPI (blue) and intestinal proliferation using EdU (green) staining; EdU was injected into flies just 3 hours before dissections. Samples were assayed on day 5 after *S. marcescens* infection at 25°C. (**B**) Representative confocal image showing JAK-STAT pathway activation in EdU-positive nuclei in an intestinal stem cell using the *stat92E*-GFP reporter line. Data are from day 5 following *S. marcescens* challenge. In a total of three experiments and 39 gut dissections, we detected 13 cells with small nuclei (DAPI) that were positive both for 10xSTAT-GFP (red) and positive for EdU staining in the region anterior to the copper cells, although no such cells were observed in 42 noninfected control guts (*P* < 0.003, Student's *t* test). EdU was injected 3 hours before dissection. (**C**) Survival curves of *S. marcescens*–infected *Drosophila* in which PP1α96A is specifically silenced in intestinal stem cells of adult flies using Esg-GAL4;tubulinGal80[ts] at 25°C. Control and *key* mutant lines are shown for comparison. *\*\*\*P* < 0.0001 (logrank test). (**D**) Analysis of gut epithelium integrity in Esg-*pp1a96A*-RNAi lines kept for 5 days after *S. marcescens* infections at 25°C (top) or under nonpathogenic conditions (bottom). Nuclei were visualized with DAPI, and actin was visualized with phalloidin (green).

sensitized background (fig. S12). In contrast, partial pathway inhibition via gut-specific overexpression of PIAS (NP1-UAS-*pias*), dominant-negative *domeless* (NP1-UAS-*domeDN*), or RNAi-mediated silencing of the *domeless* ligand, UPD (NP1-RNAi-*upd*) significantly increased the survival of *Serratia*-challenged flies (Fig. 3F). Thus, the JAK-STAT pathway activation in the gut negatively regulates survival in response to an intestinal *S. marcescens* infection.

To elucidate a possible mechanism in which JAK-STAT is involved in host defense against *S. marcescens*, we analyzed the effects of infection on gut epithelium. Infected flies exhibited massive death of intestinal epithelial cells (fig. S14A) and compensatory proliferation (fig. S14, B and C). Enhanced JAK-STAT signaling, through the use of NP1-RNAi-*pp1α96A* flies, resulted in a marked reduction in the number of large, polyploid nuclei, which signify differentiated enterocytes (*31*), after 5 days of infection (Fig. 4A). Epithelial morphology (fig. S15A) and survival on sucrose solution under nonpathogenic conditions (fig. S15B) were comparable for control, NP1-RNAi-*pp1α96A*, NP1-UAS-*pias*, and NP1-UAS-*domeDN* fly lines. We next assessed whether JAK-STAT signaling affected cellular proliferation of the epithelium. We found that DNA synthesis in epithelial cells was reduced when JAK-STAT signaling was impaired and significantly increased by silencing *pp1α96A* in the gut, both in the presence and absence of infection (Fig. 4A and fig. S16). Thus, JAK-STAT signaling enhances epithelial cell death and positively regulates compensatory proliferation of intestinal cells, also after *S. marcescens* infection.

We next examined whether the JAK-STAT pathway was affecting intestinal cell homeostasis specifically through the resident stem cell compartment (*32*). Basal intestinal stem cells (ISCs) can be distinguished from apical enterocytes on the basis of a characteristic smaller nuclear morphology (*31, 33*). By using the *stat92E*-GFP reporter line to image JAK-STAT activation, the JAK-STAT pathway was selectively induced in the ISCs but not in mature enterocytes (fig. S17). Moreover, on infection of *stat92E*-GFP flies with *S. marcescens*, we observed GFP expression also in small, 5-ethynyl-2′-deoxyuridine (EdU)–positive cells, which suggests that JAK-STAT signaling regulates ISC proliferation during *S. marcescens* infection (Fig. 4B). To definitively demonstrate that this pathway acts in gut stem cells and that this compartment controls susceptibility to *S.*

*marcescens* infections, we silenced *pp1α96A* in adult ISCs using an escargot-GAL4 driver line. Escargot is a specific marker of ISCs (*31*). ISC-specific suppression of PP1α96A resulted in early lethality in response to *S. marcescens* infection, whereas flies remained viable under nonpathogenic conditions (Fig. 4C and fig. S18). Furthermore, the guts of infected escargot-GAL4-*pp1α96A*-RNAi flies showed a phenotype similar to that obtained using the gut-specific NP1 driver, namely, severely depleted mature enterocytes (Fig. 4, A and D). Thus, our data demonstrate that JAK-STAT signaling is required for ISC homeostasis and implicates ISCs as a critical component of host defense to mucosal *S. marcescens* infections.

Our global experimental approach allows a comprehensive dissection of the biological processes that may regulate host defense against a bacterial infection at the organism level. Besides revealing previously known immune pathways, we uncovered more than 800 additional genes, many of which were of unknown function. Furthermore, our data demonstrate that host defense may involve many processes that are not limited to classical innate immune response pathways, as exemplified here by the role of the JAK-STAT pathway in the regulation of epithelial homeostasis in response to infection. In addition, we validate and map conserved candidates to intestinal cells and hemocytes, which allows us to propose a blueprint of the processes involved in host defense against *S. marcescens* infection. As all genes analyzed here are conserved during evolution, it is likely that some of the processes that are important in flies are also relevant to mammalian host defense (*34, 35*).

## References and Notes

1. B. Lemaitre, E. Nicolas, L. Michaut, J. M. Reichhart, J. A. Hoffmann, *Cell* **86**, 973 (1996).
2. D. Ferrandon, J. L. Imler, C. Hetru, J. A. Hoffmann, *Nat. Rev. Immunol.* **7**, 862 (2007).
3. B. Lemaitre, J. Hoffmann, *Annu. Rev. Immunol.* **25**, 697 (2007).
4. D. Artis, *Nat. Rev. Immunol.* **8**, 411 (2008).
5. N. T. Nehme *et al.*, *PLoS Pathog.* **3**, e173 (2007).
6. C. Flyg, K. Kenne, H. G. Boman, *J. Gen. Microbiol.* **120**, 173 (1980).
7. C. L. Kurz *et al.*, *EMBO J.* **22**, 1451 (2003).
8. Materials and methods are available as supporting material on *Science* Online.
9. G. Dietzl *et al.*, *Nature* **448**, 151 (2007).
10. L. W. Cheng *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13646 (2005).
11. S. L. Stroschein-Stevenson, E. Foley, P. H. O'Farrell, A. D. Johnson, *PLoS Biol.* **4**, e4 (2006).
12. L. M. Stuart *et al.*, *Nature* **445**, 95 (2007).
13. M. Boutros *et al.*, *Science* **303**, 832 (2004).
14. T. E. Lloyd *et al.*, *Neuron* **26**, 45 (2000).
15. M. Boutros, H. Agaisse, N. Perrimon, *Dev. Cell* **3**, 711 (2002).
16. Y. Lu, L. P. Wu, K. V. Anderson, *Genes Dev.* **15**, 104 (2001).
17. J. M. Park *et al.*, *Genes Dev.* **18**, 584 (2004).
18. N. Vodovar *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11414 (2005).
19. C. Wojcik, G. N. DeMartino, *J. Biol. Chem.* **277**, 6188 (2002).
20. C. Kocks *et al.*, *Cell* **123**, 335 (2005).
21. E. C. Lai, *Development* **131**, 965 (2004).
22. L. A. Raftery, D. J. Sutherland, *Dev. Biol.* **210**, 251 (1999).
23. H. Agaisse, N. Perrimon, *Immunol. Rev.* **198**, 72 (2004).
24. H. Agaisse, U. M. Petersen, M. Boutros, B. Mathey-Prevot, N. Perrimon, *Dev. Cell* **5**, 441 (2003).
25. N. Buchon, N. A. Broderick, M. Poidevin, S. Pradervand, B. Lemaitre, *Cell Host Microbe* **5**, 200 (2009).
26. S. R. Singh, W. Liu, X. S. Hou, *Cell Stem Cell* **1**, 191 (2007).
27. E. A. Bach *et al.*, *Gene Expr. Patterns* **7**, 323 (2007).
28. Y. C. Tsai, Y. H. Sun, *Genesis* **39**, 141 (2004).
29. A. Betz, N. Lampen, S. Martinek, M. W. Young, J. E. Darnell Jr., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9563 (2001).
30. P. Muller, D. Kuttenkeuler, V. Gesellchen, M. P. Zeidler, M. Boutros, *Nature* **436**, 871 (2005).
31. C. A. Micchelli, N. Perrimon, *Nature* **439**, 475 (2006).
32. M. Chatterjee, Y. T. Ip, *J. Cell. Physiol.* **220**, 664 (2009).
33. B. Ohlstein, A. Spradling, *Nature* **439**, 470 (2006).
34. K. Cadwell *et al.*, *Nature* **456**, 259 (2008).
35. T. Saitoh *et al.*, *Nature* **456**, 264 (2008).
36. We thank all members of our laboratories and the Vienna *Drosophila* RNAi Center for helpful discussions and technical support. We thank M. Novatchakova and M. Lafarge for expert technical help, the *Drosophila* Resource Center of the National Institute of Genetics of Japan for midgut *Gal4* driver stocks, B. Mathey-Prévot for the HML-GAL4 line, and J. Mutterer for help with confocal microscopy. This work was supported financially by the CNRS, an NIH Program grant P01 AI44220, and a DROSELEGANS grant from the Programme Microbiologie, Immunologie, et Maladies Emergentes (MIME) of the Agence Nationale de la Recherche. The D.F. laboratory is an "Équipe FRM," awarded by the Fondation pour la Recherche Médicale. J.M.P. is supported by IMBA, EuroThymaide, an Austrian Science Fund–Science Research Program (FWF-SFB) grant, an advanced European Research Council grant, and the Austrian Ministry of Science. R.M.S. and A.v.H. are supported by the Vienna Science and Technology Fund (WWTF) and the German Research Foundation (DFG) (Ha-1628/8-1). The screen was supported by Boehringer Ingelheim.

## Imaging Systems

The Leica MM AF imaging systems combine Leica's microscopy with MDS Analytical Technologies' MetaMorph software to address a range of applications in wide-field imaging in life sciences research. The tightly integrated systems are based on the ease-of-use and optical performance of the Leica microscope range, giving researchers an efficient platform for many types of imaging experiments. An integrated journaling capability allows flexible customization of the platform for virtually any imaging application.

Leica Microsystems
For information +49-(0)-6441-29-2550
www.leica-microsystems.com

## Petri Dish Filler

A compact, high-volume media preparation system provides a fast, automated method for filling Petri dishes. The Integra Mediajet is designed for high-volume microbiological media users who handle up to 20,000 plates per week. All functions can be controlled via a full-size screen. Once the dish-filling process has started the system continues the cycle through to completion without the need for input or monitoring, allowing operators to work on other tasks. The system is designed for reliable and independent operation, as the dishes are mechanically guided through the process and monitored by a set of sensors during the filling process. Typical variations in the diameter or shape of disposable plastic dishes, as well as inconsistencies in the rims or lids, are also compensated for by the system, eliminating the potential for dish jams.

Scientific Laboratory Supplies
For information +44-(0)-1159-821-111
www.scientific-labs.com

## Low Particulate Bottles

Nalgene Certified Clean PETG Containers are suitable for pharmaceutical and biotechnology applications requiring the processing and storage of bulk intermediates and critical reagents such as vaccine and protein therapeutic preparations. The containers provide excellent gas barrier properties to minimize pH shifts. The resins used to produce these containers are free of animal-derived components. The bottle with its closure system is guaranteed leakproof, making it suitable for use in transportation.

Thermo Fisher Scientific
For information 508-742-5254
www.thermofisher.com

## Ion Analysis

Two new kits are available for anion and cation analysis on the P/ACE MDQ and ProteomeLab PA 800 Systems. Capillary electrophoresis analysis of ions performed using these kits delivers efficient, rapid separations with small sample requirements and short preparation times. The kits incorporate a unique dynamic coating technology that produces consistent migration times. In drug discovery applications, the kits provide robust results for counter-ion analysis, an important part of purity determination. Compound libraries can be processed for counter-ion content even if solubility data are not available. The kits allow analysis of compounds over a wide range of polarities for applications with diverse sample matrices. These kits can also be used in the analysis of foods, beverages, and industrial products for detection and quantitation of organic acids, inorganic anions and cations, and aliphatic amines.

Beckman Coulter
For information 800-742-2345
www.beckmancoulter.com

## Cap and Tubes Strips

Masterclear Cap Strips and real-time PCR Tube Strips optimize light transmission and bring the benefits of white reflective wells to the polymerase chain reaction (PCR) tube format. The white wells of the real-time PCR Tube Strips significantly increase the fluorescent signal compared with clear or frosted wells, and reduce the interference of the thermal block. The result is improved sensitivity and reproducibility in real-time PCR experiments, which may enable the number of replicates to be reduced. The inverted dome design of Masterclear Cap Strips prevents scratching or contamination of the optical surface, while the extremely thin transmission windows ensure transmission values greater than 90 percent for wavelengths above 500 nm, with low variance. The increase in fluorescence signal can thus be distinguished from background noise at an earlier time point. The products provide particular advantages in detection of low-copy genes, when using reagents with weak fluorescent signals, and in low-volume real-time PCR assays.

Eppendorf
For information +49-40-538-01-640
www.eppendorf.com/consumables

## Adhesive Films

Pattern Adhesive Films provide high-integrity seals for 96-well and 384-well microplates. They have adhesive printed in a pattern that exactly matches the shape of a microplate well, so no adhesive comes into contact with samples or solvent. In contrast to traditional adhesive seals, piercing a Pattern Adhesive Film with a single or mutlichannel pipettor or robotic probe does not risk contaminating a sample with residual adhesive. Made from 50-μm thick PTFE film, they are tough yet easily pierced. They offer excellent optical reading properties.

Porvair Sciences
For information +44-1372-824290
www.porvair-sciences.com

Electronically submit your new product description or product literature information! Go to **www.sciencemag.org/products/newproducts.dtl** for more information.

## Pediatric Cancer Research Faculty Positions

**The Research Institute at Nationwide Children's Hospital
& The Ohio State University**

Tenure-track faculty positions are currently available at the assistant and associate professor level in the research area of molecular and cellular therapy of childhood cancer in The Research Institute at Nationwide Children's Hospital with a joint appointment in the Department of Pediatrics at the Ohio State University. Preference will be given to candidates interested in working in a highly collaborative environment with interests in childhood cancer etiology, progression and treatment. We are particularly interested in recruiting candidates with interests focused on mesenchymal stem cell biology, sarcomas, or target discovery and development of novel therapeutic agents.

The recruited candidate(s) will be provided with a generous start-up package, competitive compensation and benefits, modern laboratory space and access to state-of-the-art core facilities. The faculty member(s) will be expected to establish and maintain independent, externally funded research programs. Applicants should have at least two years, but no more than five years, of post-doctoral or equivalent experience and a strong publication record. Funded individuals will be given preference. Please send a curriculum vitae, statement of research interests and goals, and the names of three references to:

**Peter J. Houghton, Ph.D.,
The Research Institute at Nationwide Children's Hospital
WA5011, 700 Children's Drive
Columbus, OH 43205
or email: peter.houghton@nationwidechildrens.org**

*Nationwide Children's Hospital is an equal opportunity affirmative action institution that proudly values diversity. Candidates of all backgrounds are encouraged to apply.*

NATIONWIDE CHILDREN'S HOSPITAL™

3381

---

### FACULTY POSITION AT THE RIKEN BRAIN SCIENCE INSTITUTE

The Brain Science Institute (BSI) of RIKEN (Institute of Physical and Chemical Research) in Wako, Japan is seeking outstanding neuroscientists for tenure-track Team Leader positions (equivalent to a U.S. Assistant or Associate Professor) or a tenured Senior Team Leader position (equivalent to U.S. Full Professor).

The area and theme of research we wish to strengthen are broad, but we are particularly interested in investigators who study basic mechanisms of nervous system function using an interdisciplinary approach that includes some combination of genetics, electrophysiology, optical imaging and behavioral and computational neuroscience. We welcome applications from candidates working at the molecular, cellular and/or systems levels, in mammalian or other model organisms.

Candidates for the Team Leader positions should have demonstrated the strong ability to develop a significant and original research program, while candidates for the Senior Team Leader position should be one of the world leaders in his/her research field.

RIKEN BSI is an Equal Employment Opportunity Employer. Applications of women are strongly encouraged.

Applicants should submit a curriculum vitae, a summary of current and proposed research, and arrange for three letters of recommendation, all to be sent to:
**Search Committee, RIKEN Brain Science Institute
2-1 Hirosawa, Wako, Saitama 351-0198, Japan**
Fax: +81-48-467-9683, Email: **search@brain.riken.jp**
**http://www.brain.riken.jp**

---

### SCHOOL OF MEDICINE & DENTISTRY
UNIVERSITY OF ROCHESTER MEDICAL CENTER

#### Faculty Position in Neurotoxicology

Applications are invited for a tenure-track faculty position at the Associate Professor level in the Department of Environmental Medicine at the University of Rochester. We seek an outstanding individual with a strong commitment to excellence in research, scholarship, and teaching. The Department of Environmental Medicine hosts an NIEHS Environmental Health Sciences Center of Excellence, an EPA Particulate Matter Center, and an NIEHS Toxicology Graduate Training Program (http://www2.envmed.rochester.edu/envmed/index.html). This individual is expected to maintain a vigorous, creative, and independent research program, and to assume specific leadership roles within the NIEHS Center and the Department of Environmental Medicine.

We are especially interested in applicants that apply cutting-edge approaches to fundamental questions in neurotoxicology particularly those examining the relationship between environmental exposures and neurodegenerative and neurodevelopmental diseases, although exceptional candidates working in related areas are also encouraged to apply. The successful candidate will join a highly interactive faculty with research interests in neurotoxicology, pulmonary and developmental toxicology, osteotoxicology, immunotoxicology, and the molecular biology of xenobiotic receptors and transporters. Competitive salary, start-up packages and access to the many Facility Cores of the NIEHS Center and the Medical Center are provided. The University of Rochester Medical Center is undergoing a major expansion of its research facilities and infrastructure, with an emphasis on the areas of cancer, cardiovascular disease, immunology and infectious disease, musculoskeletal disease, and neuromedicine.

Applicants should send a CV, a statement of research interests including future plans, and the names of 3-5 references to: **Dr. Kerry O'Banion, Search Committee Chair, Department of Neurobiology and Anatomy, University of Rochester School of Medicine, 601 Elmwood Avenue, Box 603, Rochester, NY 14642;** (**Kerry_OBanion@urmc.rochester.edu**).

*The University of Rochester is an Equal Opportunity Employer.*

---

## Research Group Leader Position (W2)

within the

### Advanced Study Group of the Max Planck Society at the Center for Free-Electron Laser Science in Hamburg

The Advanced Study Group (ASG) of the Max Planck Society at the Center for Free Electron Laser in Hamburg (http://www.cfel.mpg.de) invites applications for the position of an Research Group Leader. The position is limited to five years with the possibility of a two year extension contingent on a positive evaluation and funding. Payment and benefits are according to W2 (Bundesbesoldungsgesetz). The starting date in 2009 is flexible. The successful candidate is expected to have a strong research profile. He/she will establish a research group of his/her own preferentially around the theory of ultrafast diffractive imaging. Possible research foci include, but are not limited to, development of new imaging algorithms or modelling microscopic dynamics of matter in FEL light.

The research group is part of the ASG and will work on the DESY campus in Hamburg. The group will be associated with the Max Planck Institute for the Physics of Complex Systems in Dresden which offers full computational and administrative support through the department Finite Systems, headed by J.M. Rost.

The group is expected to interact closely with the experimental groups within the ASG and at FLASH including access to experimental data. Contact to theory is provided through the link to mpipks in Dresden.

Applications - containing CV, publication list, statement of research plans and interests, and at least three letters of recommendation - should be sent to Gabriele Makolies, quoting 'Image09' as reference, either by electronic (makolies@pks.mpg.de) or regular mail (Max-Planck-Institut für Physik komplexer Systeme, Nöthnitzer Str. 38, 01187 Dresden, Germany). More information can be obtained directly from Prof. Dr. Jan Michael Rost (rost@pks.mpg.de).
Applications will be accepted until the position is filled; the deadline for primary consideration is Sept 1st, 2009.

The Max-Planck-Society aims to increase the number of women in scientific positions. Female candidates are therefore particularly encouraged to apply. In case of equal qualifications, candidates with disabilities will take precedence.
The Max-Planck-Institute aims to increase the number of women in scientific positions. Female candidates are therefore particularly encouraged to apply. In case of equal qualifications, candidates with disabilities will take precedence.

## CHAIRPERSON
### Department of Biochemistry and Molecular Biology

**SAINT LOUIS UNIVERSITY**

Saint Louis University, a Catholic, Jesuit institution dedicated to student learning, research, healthcare, and service has initiated a national search for the Chairperson of the Edward A. Doisy Department of Biochemistry and Molecular Biology. The department has a strong tradition of excellence in research and education. It was founded in 1923 by Dr. Edward A. Doisy who received a Nobel Prize for his research on Vitamin K. The department occupies ample space in the recently built Doisy Research Center (**http://researchbuilding.slu.edu/**), with outstanding research equipment and core facilities. Departmental faculty are consistently funded in the areas of protein structure and function, macromolecular interactions, gene regulation, signal transduction, protein trafficking and turnover, cardiovascular research, and genetic disorders. Research strengths in other departments include neuroscience and aging, molecular virology, immunology and vaccine development, and liver disease. The School of Medicine is seeking a highly motivated academic leader who will cultivate an environment of excellence in research and teaching, and will foster an environment of collegiality and collaboration with other departments and schools. The successful applicant will provide strong leadership for all research and academic activities of the department and will have outstanding skills in communication and mentoring. She/he will have a Ph.D., M.D./Ph.D. or M.D. degree, a distinguished record of research productivity, teaching excellence, and leadership skills in academic program development.

All applications must be made at: **http://jobs.slu.edu** and must include a curriculum vitae and cover letter. Letters of nomination(s) and curriculum vitae may be sent to:

**William S. M. Wold, Ph.D.**
**Chairperson, Biochemistry Chairperson Search Committee**
**Saint Louis University School of Medicine**
**1402 South Grand Boulevard, M268**
**St. Louis, MO 63104**
**Email: woldws@slu.edu**

*Saint Louis University is an Affirmative Action/Equal Opportunity Employer and encourages nominations and applications from women and minorities.*

## BIOCHEMISTRY FACULTY POSITION

Kansas City University of Medicine and Biosciences invites applications from outstanding candidates for appointment in the Department of Biochemistry at the rank of Assistant or Associate Professor. As we continue to expand our research capacity, we seek candidates with research interests in the chronic diseases of aging, particularly diseases associated with protein misfolding; applicants who can produce new knowledge in the areas of chemical neurobiology using innovative techniques are especially encouraged to apply. The successful applicant will have a Ph.D. (or equivalent doctorate), a record of scholarly publications and funding, be willing to mentor graduate and medical students and to significantly contribute to a novel instructional curriculum. For additional information, contact **Norbert W. Seidler, Ph.D., Chair, Department of Biochemistry, 1-800-234-4847, ext. 2207 or 816-283-2207, nseidler@kcumb.edu**.

KCUMB, a growing institution with an emerging emphasis on research, recently completed a 46,000 sq ft building for bioscience research. The College of Osteopathic Medicine is Missouri's largest medical school and strives to hire outstanding faculty and staff to provide an exemplary medical education for approx 1,000 osteopathic medical and biosciences graduate students. Excellent pay is complemented with an exceptional benefits package. Located in the Northeast historic district of Kansas City, MO, near downtown and collaborating institutions, KCUMB is one of eight original key stakeholder institutions in the Kansas City Area Life Sciences Institute. **www.kclifesciences.org**.

Excellent pay is complemented with an exceptional benefits package. Salary and rank will be commensurate with experience and qualifications. To apply, submit a cover letter for job #09-17 (include salary range requirements), curriculum vitae, a statement of teaching philosophy and research interests, strengths for the position, and contact information for three references to: **Nicole Torgerson, Recruiter, 1750 Independence Ave., Kansas City, MO 64106-1453, 1-800-234-4847, ext. 2229 or 816-283-2229;** or **e-mail: employment@kcumb.edu** (Word or PDF format only please), or **fax 816-283-2285**.

*Pre-employment drug screen and background check required. EOE.*

**www.kcumb.edu**

**KANSAS CITY UNIVERSITY MEDICINE & BIOSCIENCES**

## Leibniz-Institute of Plant Biochemistry

The Department of Molecular Signal Processing at the Leibniz-Institute of Plant Biochemistry (IPB) in Halle (Saale), Germany, invites applications for three positions of

# Research Group Leader

The focus of the department (http://www.ipb-halle.de/en/) is to understand at the molecular level how plant systems monitor and respond to environmental challenge by adapting metabolism and growth. This is a search for outstanding candidates to establish research in the integrated program areas of nutrient sensing (P, Fe, N), defense metabolism (glucosinolates) and hormone biology (auxin, jasmonates), with particular emphasis on plant-rhizosphere interactions using plant model systems.

Candidates are expected to develop original research on an internationally competitive level. Successful applicants will have a command of modern experimental and bioinformatics-related skills as well as a strong track record of high-quality publications that reflect their creativity and impact. Excellent writing and communication skills for securing extramural funds and disseminating research are a prerequisite. For each position, funds are available for a postdoctoral scientist, a PhD student and a technician, together with funds for consumables and equipment. The appointment will be initially for 5 years with the possibility of extension. Compensation will be according to TV-L and professional experience. Fluency in English is required.

The IPB, located on the *weinberg-campus* of the Martin-Luther-University Halle-Wittenberg, is an international research institute with about 200 employees and provides state of the art facilities for research in natural products chemistry/metabolomics, plant biochemistry/proteomics, and cell biology/imaging.

Applications from qualified researchers of any nationality are welcomed and should include a short description of research accomplishments and future plans, *curriculum vitae,* publication list, and contact information for three referees. Applications should be sent by **September 15, 2009** to:

**Steffen Abel, Leibniz-Institute of Plant Biochemistry**
**Weinberg 3, D-06120 Halle (Saale), Germany**
**or via email to: sabel@ipb-halle.de**

## Director, Stem Cell Research

The Department of Surgery at the University of Arizona is seeking a Stem Cell Biologist who will serve as Director, Stem Cell Research, for the Department's Institute for Cellular Transplantation, currently focusing on islet cell transplant both clinically and from a basic science perspective. The selected candidate will direct research in the stem/progenitor cell biology of pancreatic Beta Cells.

Requirements include PhD or MD/PhD in molecular or cellular biology. Experience preferred in approaches to study intracellular and intercellular signal transduction pathways and the role of the microenvironment in regulating cell fate.

The tenure-track faculty position will be filled at the Assistant Professor or Associate Professor rank and includes a competitive salary, start-up funds and new laboratory space in the Arizona Health Sciences Center.

Existing Surgery Department strengths are in diabetes, therapeutic uses of islet cell transplantation, in oncology and cancer stem cells. The successful candidate will develop an independent sustainable research program and contribute to departmental teaching in the Medical Sciences Graduate Program. Ample opportunity for collaboration exists with Colleagues in Molecular and Cellular Biology, Physiology and Bioengineering.

Apply on-line at **www.hr.arizona.edu** Job posting #43333. Position is opened until filled.

Send CV and Letter of interest, description of past research accomplishments and future goals in research, and names and contact information for two references to:

**Dr. Rainer W.G. Gruessner**
**Department of Surgery**
**University of Arizona**
**1501 N. Campbell Avenue, Rm. 4410**
**POBox 245066**
**Tucson, AZ 85724**
**rgruessner@surgery.arizona.edu**

*The UA is an EEO/AA Employer.*

## Postdoctoral Fellowship Awards in the Early Detection of Cancer

**American Cancer Society®**

**CANARY FOUNDATION**
*Stopping Cancer Early - The Best Possible Investment*

The Canary Foundation, in partnership with the American Cancer Society, is extending its postdoctoral fellowship program focused on **studies towards development of strategies for the early detection of cancer**. Research should be directed at new approaches to improve clinical methods for the detection of primary tumors and/or metastases, including but not limited to research focused in the following areas:

- Minimally invasive strategies for early detection
- Biomarker identification or discovery
- Imaging, including novel molecular imaging strategies
- Segmentation of early stage disease according to prognosis

Awards will be 3 years in duration with progressive stipends of $44,000, $46,000, and $48,000 per year, plus $4,000 per year for fellowship allowance. Applications will only be accepted from scientists who, at the time of application, have had no more than 2 years of research experience beyond their terminal degree (MD or PhD). Applicants must be US citizens or permanent residents working with an accomplished mentor at a non-profit institution. Awardees will be asked to attend the Canary Foundation Early Detection Symposium May 25-27, 2010 and to contribute to the online Canary Journal project. We anticipate awarding up to 4 fellowships.

**Deadline:** Complete application: **October 15, 2009**. For additional information regarding program policies or to obtain an application, please refer to the ACS website: **www.cancer.org/research**. To learn about the Canary Foundation, please visit **www.canaryfoundation.org**. For specific inquiries, contact **Michael H. Melner, PhD, Scientific Program Director at 404-327-6528** (**michael.melner@cancer.org**).

---

**ASSISTANT PROFESSOR OF BIOCHEMISTRY**
**UNIVERSITY OF WASHINGTON, SEATTLE**

**The Department of Biochemistry at the University of Washington School of Medicine** invites applications for a tenure-track position at the Assistant Professor level. We seek creative scientists whose research focuses on molecular aspects of important biological problems. Our department is also strongly committed to teaching at the undergraduate, graduate, and medical school levels.

Applicants should have a Ph.D. in biochemistry, cellular or molecular or developmental biology, structural biology or biophysics, or a related field, or an M.D. Submit a curriculum vitae, research prospectus, and reprints or preprints as PDFs to **bcsearch@u.washington.edu** by September 30, 2009.

Three letters of recommendation should be emailed separately or mailed to **Alan Weiner, ZymoGenetics Chair, Department of Biochemistry, University of Washington, Seattle, WA 98195-7350. Voice: 206-543-1768.**

*The University of Washington is an Affirmative Action/Equal Opportunity Employer.*

# Everyday needs. Everyday answers.™