TEAM tAO

12 August 2005

Science

Vol. 309 No. 5737
Pages 973–1136 $10

Dealing with DISASTERS

Science

Special Issue

Dealing with Disasters

A month after the December 2004 tsunami, people use a makeshift ferry to cross a river in Lokinga, Indonesia, as soldiers work to reconstruct a bridge. Recent devastating disasters have sparked renewed efforts to prepare for and speed recovery from such unpredictable threats. [Photo: Jewel Samad/AFP/Getty Images]

Volume 309
12 August 2005
Number 5737

For related online content in Science's Next Wave, see page 983 or go to www.sciencemag.org/sciext/disasters/

Departments

News of the Week

News Focus

1002

1027

Letters

Books *et al.*

Policy Forum

Perspectives

Contents continued

Science

Science Express www.sciencexpress.org

Plant Science: The mRNA of the *Arabidopsis* Gene *FT* Moves from Leaf to Shoot Apex and Induces Flowering

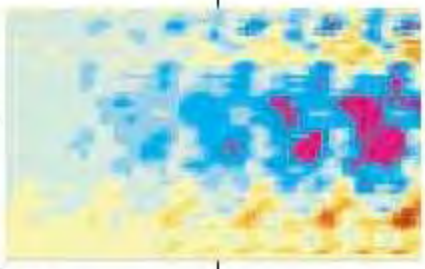
T. Huang, H. Böhlenius, S. Eriksson, F. Parcy, O. Nilsson

The long-sought "florigen" that moves from leaf to shoot and induces flowering as days lengthen is the messenger RNA for the *FLOWERING LOCUS T* gene *FT*. *related Perspective page 1024; Research Articles pages 1052 and 1056*

Atmospheric Science: The Effect of Diurnal Correction on Satellite-Derived Lower Tropospheric Temperature

C. A. Mears and F. J. Wentz

After modification of an erroneous diurnal correction, a reconstruction of recent atmospheric warming of the lower troposphere from satellite data now agrees with that inferred from measurements at the surface. *related Science Express Reports by B. D. Santer et al. and S. Sherwood et al.*



Atmospheric Science: Amplification of Surface Temperature Trends and Variability in the Tropical Atmosphere

B. D. Santer et al.

Results of modeling recent temperature changes in the tropical troposphere agree with satellite data that indicate more warming than earlier reconstructions. *related Science Express Reports by C. A. Mears and F. J. Wentz; S. Sherwood et al.*

Atmospheric Science: Radiosonde Daytime Biases and Late–20th Century Warming

S. Sherwood, J. Lanzante, C. Meyer

Temperature measurements by weather balloons in the troposphere failed to reveal the extent of warming because of an uncorrected artifact in new instrumentation. *related Science Express Reports by C. A. Mears and F. J. Wentz; B. D. Santer et al.*

Technical Comment Abstracts

1017 **Ocean Science**

Comment on "Molybdenum Isotope Evidence for Widespread Anoxia in Mid-Proterozoic Oceans"

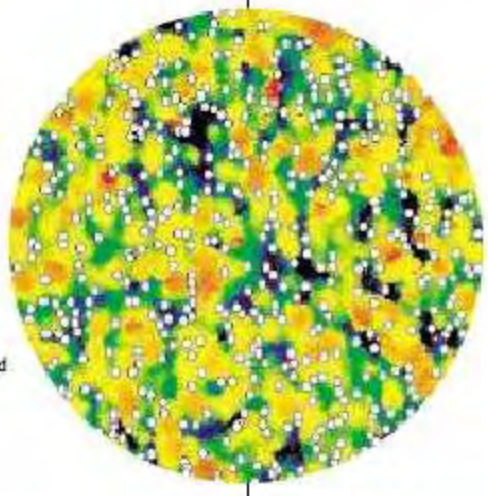
H.-F. Ling, J.-F. Gao, K.-D. Zhao, S.-Y. Jiang, D.-S. Ma

full text at www.sciencemag.org/cgi/content/full/309/5737/1017c

Response to Comment on "Molybdenum Isotope Evidence for Widespread Anoxia in Mid-Proterozoic Oceans"

A. D. Anbar, G. L. Arnold, T. W. Lyons, J. Barling

full text at www.sciencemag.org/cgi/content/full/309/5737/1017d



Brevia

1047 **Ecology:** Soil Invertebrates Disrupt Carbon Flow Through Fungal Networks

D. Johnson et al.

The massive flux of carbon through plants to soil via symbiotic mycorrhizal fungi can be reduced 32 percent by tiny soil insects that feed on the fungi.

Research Articles

1048 **Physics:** Atomic-Scale Sources and Mechanism of Nanoscale Electronic Disorder in $Bi_2Sr_2CaCu_2O_{8+\delta}$

K. McElroy, J. Lee, J. A. Slezak, D.-H. Lee, H. Eisaki, S. Uchida, J. C. Davis

Dopant atoms in a high-temperature superconductor are shown by scanning probe microscopy to be close to areas of electronic disorder. *related News story page 1001*

1001 & 1048

Plant Science

1052 FD, a bZIP Protein Mediating Signals from the Floral Pathway Integrator FT at the Shoot Apex

M. Abe et al.

1056 Integration of Spatial and Temporal Information During Floral Induction in *Arabidopsis*

P. A. Wigge et al.

Two nuclear genes, one activated in the leaf and one in the shoot, work together to determine the time and location of flowering. *related Perspective page 1024; Science Express Research Article by T. Huang et al.*

Reports

1059 **Materials Science:** Origin of Brittle Cleavage in Iridium

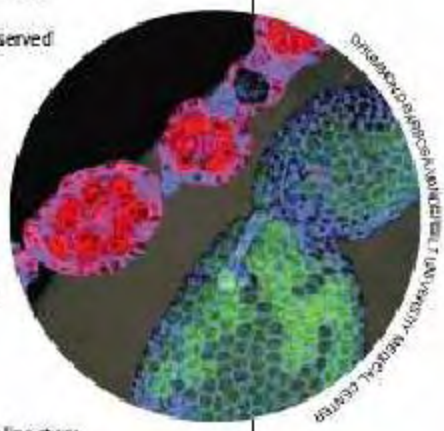
M. J. Cawkwell, D. Nguyen-Manh, C. Woodward, D. G. Pettifor, V. Vitek

Iridium's characteristic brittle cleavage, not seen in any other metal with its structure, is caused by the formation of two types of dislocations and rapid interchange between them.

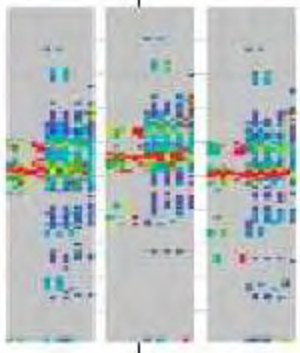
Contents continued

Science

REPORTS CONTINUED



1071



1078

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals Mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2005 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership and subscription (51 issues): $135 ($74 allocated to subscription) Domestic institutional subscription (51 issues): $932. Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air mail: all to ry): 80s First class, airmail, student, and retired rates on request. Canadian rates with GST available upon request, GST # 125439812Z. Publications Mail Agreement Number 1069634. Printed in the U.S.A.

Change of address allow 4 weeks, giving old and new addresses and 8-digit account number. Postmaster: Send change of address to Science, P.O. Box 1811, Danbury, CT 06813–1811. Single copy sales: $10.00 per issue prepaid and sales outside postage; bulk rates on request. Authorization to photocopy material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $15.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for Science is 0036-8075/03 $15.00. Science is indexed in the Reader's Guide to Periodical Literature and in several specialized indexes.

Contents continued

sciencenow www.sciencenow.org Daily News Coverage

What's Your Dog Thinking?
You may not be able to tell, but other canines seem to get the picture.

Brewing a Better Killer
DNA repeats explain how yeast evade drugs and the immune system.

How to Sell Humvees to Men
Attacking their machismo makes men more supportive of war, more homophobic, and more willing to shell out for that SUV.

Opportunities in researching disasters.

science's next wave www.nextwave.org Career Resources for Young Scientists

Related Dealing with Disasters section page 1029

- **Global: Careers in the Science of Disasters—Feature Index** *A. Forde*
  Next Wave investigates scientists who use their expertise to help predict, prevent, and deal with disasters.
- **Global/US: Little Movement in Earthquake Science Careers** *J. Kling*
  Earthquake science careers have never been high demand, but opportunities do exist.
- **Global/Europe: Monitoring the Pulse of the Mount Vesuvius** *E. Pain*
  Massimo Orazi satisfies his appetite for technical advances at the Vesuvius Observatory.
- **Global/Europe: Learning from Disasters** *A. Forde*
  Ilan Kelman talks to Next Wave about his broad interest in the science of disasters.
- **Global/GrantsNet: Student and Postdoctoral Funding in the Science of Disasters** *Edited by S. Martin*
  GrantsNet offers a sampling of funding opportunities in disciplines related to the science of disasters.
- **Career Development Center: Mitigating Disasters** *R. Amette*
  A disaster epidemiologist ensures that the health needs in disaster-stricken areas are assessed and met.

science's sage ke www.sageke.org Science of Aging Knowledge Environment

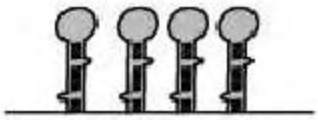
Genetically Altered Mice: BRI-Aβ42 Transgenic Mice *Q. Guo*
Mice that produce human Aβ 1–42 in the absence of human amyloid precursor protein production develop amyloid-related pathologies.

Genetically Altered Mice: BRI-Aβ40 Transgenic Mice *Q. Guo*
Mice that produce human Aβ 1–40 in the absence of human amyloid precursor protein production do not develop amyloid-related pathologies.

News Focus: Partly Cloudy *R. J. Davenport*
Mice with Alzheimer's disease–linked gene variants suffer deteriorating retinas.

News Focus: Up to Speed *M. Leslie*
Amphetamines spare mice from Parkinson's disease–like symptoms.

Mousing for clues about fading vision.



Primary microRNA transcripts.

science's stke www.stke.org Signal Transduction Knowledge Environment

Perspective: Slowing Down the Ras Lane—miRNAs as Tumor Suppressors? *J. P. Morris IV and M. T. McManus*
Do microRNAs act as tumor suppressors or oncogenes by regulating the expression of particular targets?

Teaching Resource: Regulation of Ion Channels by G Proteins *M. Diversé-Pierluissi*
Prepare a graduate-level class covering the regulation of calcium channels through G-protein signaling.

Separate individual or institutional subscriptions to these products may be required for full-text access.

Dopant Distribution and Superconductivity

On the microscopic scale, the electronic structure of the high-temperature (high-T_c) superconducting cuprates is inhomogeneous, and this nanoscale electronic disorder could be caused by a random distribution of dopant atoms. However, the identification of the dopant atoms in the material, and quantifying what influence they have on the electronic structure has been difficult to realize. **McElroy *et al.*** (p. 1048; see the news story by **Cho**) used scanning probe microscopy to image the dopant atom locations and probe the atomic-scale electronic structure simultaneously in the high-T_c superconductor $Bi_2Sr_2CaCu_2O_{8+x}$. Correlating the dopant atom distribution with the electronic properties may provide a clearer understanding of not only the cuprates but also other doped complex materials.

More Is...Less?

The abundance of photosynthetic organisms in much of the upper ocean is thought to be limited by the amount of the essential nutrient phosphorus (P) that is available. Thus, the number of primary producers in a P-depleted ocean region would increase if this nutrient was added. **Thingstad *et al.*** (p. 1068) performed such an experiment on a 16-square-kilometer area of the Mediterranean Sea, where productivity is very low and P is the limiting nutrient. The chlorophyll content of the waters actually decreased, opposite what would be expected, and the abundance of copepod eggs, ciliate biomass, and bacterial production all increased. The authors discuss several possible reasons for this response and how the effects of P limitation might differ with respect to season and to groups of organisms.

Sum Information Recorded

The Inka Empire was the largest pre-Columbian empire in the New World, yet apparently it lacked a written language. What is preserved are khipu, groups of intricately knotted colored strings that are thought to be used for accounting or record keeping, but deciphering their meaning, relationship, or significance has been problematic. **Urton and Brezine** (p. 1065; see the news story by **Mann**), working with seven khipu that do have some contextual information so that they can be treated as a group, show that successive khipu record summations of other ones. It appears that khipu were used to pass accounting information upward through the Inka bureaucracy.

A Time and a Place for a Flower

Early plant responses to spring environmental signals, such as day length or vernalization, occur in the leaves, but the flowers form at a meristem that shifts from vegetative to floral growth. The mysterious mobile mediator between these two sites has been called "florigen." **Wigge *et al.*** (p. 1056) and **Abe *et al.*** (p. 1052) investigated the integration of floral initiation signals in *Arabidopsis* (see the Perspective by **Blázquez**). *Flowering locus T* (*FT*) is expressed in leaves in response to environmental changes conducive to flowering. Meanwhile, the meristem is primed to be ready for floral production by expression of a transcription factor FD. Together, FT and FD activate floral identity genes. Whether FT actually is the mobile mediator remains unclear, but these two genes together do integrate the "when" and the "where" of the flowering response.

Oscillating Flashes

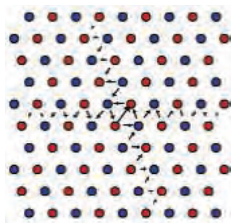
Proteins sample an enormous conformational space as they fold and unfold, and traditional measurements reveal limited details of the process because each molecule may follow a slightly different path at a slightly different time. Fluorescence measurements with single-molecule resolution can overcome this blurring effect and trace individual pathways. **Baldini *et al.*** (p. 1096) examined a green fluorescent protein mutant suspended in a gel during unfolding. Just prior to unfolding, the protein chromophore oscillates with remarkable regularity between two states, a blue-fluorescing neutral state and green-fluorescing anionic state. The oscillatory frequencies fall in the 400- to 1000-Hertz range, and the process can be driven by applied resonant electric or acoustic fields. The molecular mechanism underlying this process remains elusive.

Choosing Your Fate

During differentiation, cell lineages must choose between different alternate fates. **Hong *et al.*** (p. 1074) provide evidence that a protein known as TAZ (for transcriptional coactivator with PDZ-binding motif) is a key regulator that helps determine the fate of mesenchymal stem cells that can differentiate into osteoblasts or adipocytes. TAZ contains a protein interaction domain that binds to Pro-Pro-X-Tyr motifs (where X represents any amino acid). Two transcription factors that control differentiation of mesencymal stem cells, Runx2 and PPARγ, contain such a motif in their activation domains. In tissue culture and in zebrafish embryos, TAZ promotes formation of osteoblasts (by cooperating with Runx2) and inhibits differentiation of adipocytes (by antagonizing the effects of PPARγ).

Bend and Snap

Iridium is the only face-centered cubic metal that undergoes brittle cleavage after deformation. Using simulations, **Cawkwell *et al.*** (p. 1059) account for this behavior as arising from the interchange of two dislocation types within the material. The cores of screw dislocations in iridium can be either planar (glissile, and moving readily) or be distributed on two planes, causing them to be nonplanar (sessile, and moving only under high stress). The athermal transformation of one defect type into the other leads to a rapid increase in the dislocation density. This process in turn leads to fast strain hardening, which eventually causes the flow stress to rise fast enough to cause brittle failure.



CONTINUED ON PAGE 987

CREDITS (TOP TO BOTTOM): WIGGE *ET AL.*; CAWKWELL *ET AL.*

Published by AAAS

Containing a Potential Pandemic

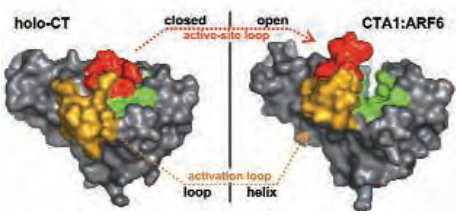
In the 20th century, there were three influenza pandemics. Currently, the world is threatened by avian influenza in Southeast Asia and may be only a reassortment or mutation event away from another pandemic. However, there is a good chance of preventing the spread of any emergent influenza strain at the source through good surveillance and the aggressive use of influenza antiviral agents, quarantine, and vaccines. In a detailed epidemic simulation model for a Southeast Asian population, **Longini *et al.*** (p. 1083, published online 4 August 2005) analyze possible strategies for containing a newly emergent influenza strain and show that such a strain should be containable at the source under a broad set of potential conditions.

How Rice Releases Methane

Rice agriculture is possibly the biggest source of anthropogenic methane: Rice paddies cover about 130 million hectares of the earth's surface, of which almost 90% are in Asia, and emit 50 to 100 million metric tons of methane a year. Most of this methane is derived from rice photosynthates excreted into the rhizosphere. **Lu and Conrad** (p. 1088) used pulse-labeling of rice plants with $^{13}CO_2$ followed by in situ stable isotope probing of rhizospheric archaeal RNA to show that a group of methanogenic archaea, the so-called Rice Cluster I, of which no isolates exist so far, is responsible for this methane production from the degradation of photosynthates.

Scrutinizing Cholera Toxin

Cholera toxin catalyzes reactions that lead to the devastating diarrhea characteristic of the disease. The toxin is activated by a family of human G proteins, adenosine diphosphate–ribosylation factors (ARFs), which normally act as molecular switches through binding effector proteins in eukaryotic cells. **O'Neal *et al.*** (p. 1093) now report high-resolution structures of the catalytic cholera toxin A1 subunit (CTA1) bound to ARF–guanosine triphosphate (GTP), with and without substrate bound. Although cholera toxin is not structurally similar to human protein partners of ARF, the toxin:ARF-GTP interface mimics ARF-GTP recognition of human effector proteins. The binding causes conformational changes that open the CTA1 active site to substrate access.



CREDIT: O'NEAL *ET AL.*

Planktonic Pathogen Genome

Given the incredible number of viruses present in the ocean, there are surprisingly few marine viral genomes known. **Wilson *et al.*** (p. 1090) provide a complete genome sequence, annotated via a microarray analysis, of a *Coccolithovirus* pathogen of the ubiquitous and globally important phytoplankton *Emiliania huxleyi*. This huge viral genome contains a family of noncoding repeats and a viral RNA polymerase gene that might function together as transcription machinery. The genome also appears to contain an apoptosis activation system, which may be pivotal in understanding the bloom behavior of the host alga. The majority of the genes in the virus are transcribed and the virus can act as a vehicle for horizontal gene transfer within this species of coccolithophorids; indeed, *E. huxleyi* is one of the fastest evolving species of phytoplankton known.

Network Analysis of Cell Regulation

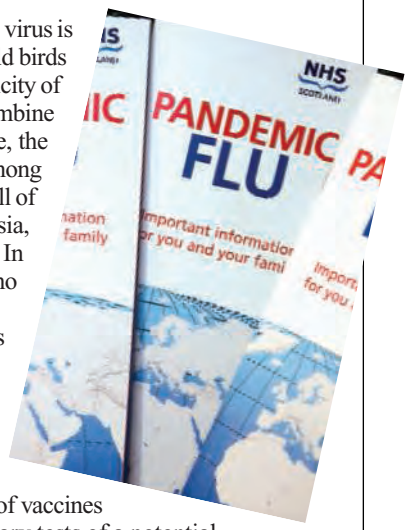
Fuller understanding of cellular regulation requires analysis of the interactions of multiple signaling pathways in the complex networks that control cellular functions. **Ma'ayan *et al.*** (p. 1078) analyze the network properties of 545 components that undergo 1259 interactions underlying signaling and cell regulation in hippocampal neurons. The presence of regulatory motifs and other characteristics begin to reveal how cells may process information to allow, for example, transient or stable changes in cell function.

Published by AAAS

Heading Off an Influenza Pandemic

Continual news on the outbreak status of the H5N1 subtype of influenza A virus in Southeast Asia points to one of the greatest challenges facing 21st-century society: the prediction and management of disasters. Hundreds of thousands die from influenza annually, with widespread and often devastating pandemics occurring episodically. The last flu pandemic occurred in 1968. Are we better able to mitigate the effect of a new pandemic than we were 37 years ago? Advances in science, vaccine strategies, and antiviral drugs provide this potential, but whether these can be applied in the short term in an effective global policy is not guaranteed.

The continual threat of influenza A viruses such as avian H5N1 lies in their basic biology. The virus is easily transmitted and can be highly virulent. It is present at high frequencies in reservoirs of wild birds that can infect domestic animals, including horses, pigs, and poultry. Add to that the genetic plasticity of the virus, with high rates of mutation and a ready capacity for reassortment that allows it to combine with other strains to produce new and sometimes highly pathogenic variants. Most worrisome, the reassortment of human and bird strains could result in a pandemic virus that is transmissible among humans. A highly pathogenic avian H5N1 virus first appeared in Hong Kong in 1997. A mass cull of chickens alleviated the problem locally, but H5N1 viruses continued to circulate in birds in Asia, most recently among migratory species that could theoretically carry the virus for long distances. In 2003, H5N1 re-emerged in humans, causing almost 60 deaths in Asia to date, although there is no convincing evidence that the virus has evolved human-to-human transmission.



How might we prevent and manage a future influenza pandemic? The most obvious requirement is a rapid and expansive influenza surveillance and response network. A permanent global task force has been proposed to perform this role,* and we strongly endorse this idea. However, such a task force will only be successful if national governments release data promptly and adhere to control measures should a pandemic arise. Such surveillance activity also needs to include humans, domestic animals, and wild birds. Second, we must develop further, effective intervention strategies to reduce transmission and disease. The development of vaccines against H5N1 strains, and ultimately against all subtypes, is a clear priority. Recent preliminary tests of a potential vaccine are encouraging. But using traditional approaches, fewer than 500 million people could currently be vaccinated with a two-dose monovalent pandemic influenza vaccine. New vaccine methodologies are in reach, but international agreements on production, intellectual property, distribution, and administration need to be aggressively pursued. Antiviral drug stockpiles are equally limited. Thus, because rapid global distribution networks of vaccines and antiviral agents have yet to be established, it is essential that logistical simulations be conducted to determine their possible limitations. Third, we need epidemiological models, to explore the spread and impact of potential influenza pandemics in the face of realistic control measures and how to manage pandemics when they arise. Two recent reports suggest that antiviral-based containment policies could be an effective strategy,† although this is contingent on a rapid, coordinated response to the emergence of a pandemic.

We must also develop strategies to reduce the probability of pandemics. This will require a multitude of basic scientific information, including the probability and mechanism of reassortment; a measure of the exposure rates of influenza viruses at the human/animal interface; and, most critically, an understanding of how avian viruses evolve to develop sustained transmission networks in humans. It is therefore essential to conduct a global surveillance of genetic diversity in avian influenza viruses, sequencing complete genomes from these and mammalian strains to explore the polygenic nature of host adaptation. We also need to determine the extent of clonal genetic variation within individual hosts, because consensus sequences invariably hide strains with varying phenotypic properties. These data would also provide perspective on the evolutionary dynamics of viral pathogens at different spatial scales. Although a unified political effort is essential to avert or mitigate a major influenza pandemic, it must proceed in parallel with advances in basic science.

The potential for avian H5N1 to cause a global human pandemic is presently uncertain because it cannot be predicted with current data. However, if an H5N1 pandemic does not emerge in the near term, the political will to continue the global preparations necessary for a future pandemic may falter. We cannot afford such a misstep.

Edward C. Holmes, Jeffery K. Taubenberger, Bryan T. Grenfell

Edward C. Holmes and Bryan T. Grenfell are at the Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, Mueller Laboratory, University Park, PA 16802, USA. Jeffery K. Taubenberger is in the Department of Molecular Pathology, Armed Forces Institute of Pathology, Rockville, MD 20850, USA.

*R. A. M. Fouchier, T. Kuiken, G. Rimmelzwaan, A. Osterhaus, *Nature* **435**, 419 (2005); S. P. Layne *et al.*, *Science* **293**, 1729 (2001).

†I. M. Longini Jr. *et al.*, *Science* **309**, 1083 (2005); N. M. Ferguson *et al.*, *Nature*, 3 August 2005 (10.1038/nature04017).

10.1126/science.1117128

Published by AAAS

edited by Gilbert Chin

ECOLOGY/EVOLUTION

Hidden Long-Term Consequences

Burned (inset) and unburned (right) montane cloud forest.

Fires are becoming increasingly frequent in wet tropical forests as a result of human land use and other associated disturbances. In addition to their visibly destructive effects on aboveground biomass, tropical forest fires can smoulder underground for a long time. Thus, fire has a potential to alter soil properties directly, especially the concentration and spatial distribution of nutrients—both of which have ramifications for the subsequent ecological dynamics of forests.

Blair has examined the effects of underground fires on the spatial patterns of soil constituents in a lowland wet forest in Nicaragua. Fire altered the spatial scale of nutrient distribution, generally reducing the patch size for key nutrients such as carbon, nitrogen, phosphorus, and potassium, with as yet unknown consequences for belowground competition between plants. Asbjornsen *et al.* describe the effects of fire on plant biomass in montane cloud forests in Mexico, a habitat type in which forest fires have been documented only in the past few years. Here, the belowground effects were substantial: Deep ground fires occurring in 1997–1998 resulted in a 50 to 75% reduction in live root biomass, as well as >80% reduction in aboveground biomass. Given the time scale of forest dynamics, the longer-term effects of these disturbances will unfold over decades. — AMS

J. Trop. Ecol. **21**, 435; 427 (2005).

MATERIALS SCIENCE

Stronger Steel

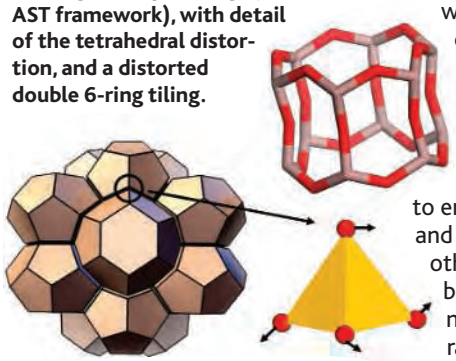
Ferritic and martensitic steels are preferred structural materials for use at elevated temperatures in power plants. Their major advantage is good thermal behavior relative to other elevated-temperature alloys, but they suffer from not being strong enough at high temperature. Niobium, vanadium, and nitrogen have been added to push upward this maximum-use temperature, but alloying may have reached its limit for enhancing these steels. The development of alloys that are strengthened through the dispersion of oxide particles requires expensive manufacturing techniques. Under normal processing conditions, commercial steel alloys develop large metal (M) carbon precipitates or MX particles, where X is Nb or V. Small MX precipitates confer higher-temperature stability, and Klueh *et al.* have developed methods to increase their density. High-temperature rolling was used to create dislocations in the alloys that acted as nucleation sites for the MX particles, increasing the nanoparticle density by three orders of magnitude. Tests on treated commercial alloys showed increases in yield stress, particularly above 620°C, which is the upper use temperature for the untreated alloys. — MSL

Scripta Materialia **53**, 275 (2005).

CHEMISTRY

Avoiding Distortions

The formation of silica zeolites usually requires some structural distortion, but what has been observed experimentally is that the bending within the SiO_4 tetrahedra is minor and occurs mainly along the bonds connecting them. Zwijnenburg *et al.* performed a computational study of known zeolitic frameworks, as well as hypothetical frameworks built up of three-connected polyhedral tiles (simple tilings such as double 4-rings), whose packing can be used to represent known frameworks. They found that the experimental frameworks had minimal tetrahedral distortion (less than 2.5×10^{-2} Å^2), but that hypothetical frameworks, even ones with enthalpies similar to those of experimental frameworks, had an order of magnitude or more tetrahedral distortion. They conclude that although a few of the hypothetical frameworks may be realizable, most of the more distorted ones will likely elude hydrothermal synthesis. — PDS



Packing of simple tilings (the AST framework), with detail of the tetrahedral distortion, and a distorted double 6-ring tiling.

J. Phys. Chem. B 10.1021/jp0531309 (2005).

PSYCHOLOGY

Of Morals and Mores

What is it that makes moral beliefs nonnegotiable? Such beliefs are thought to transcend cultural variation—in short, to be universal—and they are often associated with intense emotion, as are strongly held attitudes. Across a range of situations, however, Skitka *et al.* find evidence consistent with their proposal that attitudes and moral convictions differ. When questioned about their social distance preferences, where in-laws were proximal and public officials were distal, study participants were less tolerant of social relations with people whose convictions disagreed with their own than when the discordant positions on issues were regarded merely as a clash of attitudes. This abstract type of preference could in fact be converted into a simple physical measure of how close to another person with known similar or dissimilar convictions a participant chose to sit. Furthermore, grouped participants demonstrated a greater willingness to engage in discussion and negotiation with others when opposing beliefs arose from nonmoral attitudes rather than convictions,

CREDITS: (TOP) HEIDI ASBJORNSEN; (BOTTOM) ZWIJNENBURG *ET AL., J. PHYS. CHEM. B* 10.1021/JP0531309 (2005)

Published by AAAS

in line with recent work suggesting distinct emotional versus cognitive neural substrates for intuitive versus reasoned social appraisals. — GJC

J. Pers. Soc. Psych. **88**, 895 (2005).

GEOPHYSICS

Gradation of Fabric

The solid inner core of Earth grew larger over time as the liquid iron outer core crystallized. This process released latent heat that helps drives convection in the liquid outer core, producing Earth's magnetic field; the inner core now has a radius of about 1200 km. Recent seismic observations have started to reveal details of the inner parts of the inner core. Initial results showed that the core has a distinct and organized crystal fabric. More recently, it has been suggested that there may be a boundary deep in the inner core at a radius of about 300 km, which may indicate its episodic growth. Cormier and Stroujkova searched for additional evidence of such a structure in a series of waveforms of seismic waves passing through the inner core at various angles (thus sampling it through different depths). The data suggest, albeit not conclusively, more subtle variations in the fabric of the inner core with depth, consistent with an increase in organized crystal size with depth, instead of a sharp transition. — BH

Earth Planet. Sci. Lett. **236**, 96 (2005).

CHEMISTRY

ROMP with Restraint

Ring-opening metathesis polymerization (ROMP) yields a versatile range of linear polymers from cyclic olefin starting materials. The reaction is driven by relief of the bonding strain inherent to the geometry, as a metal catalyst pries open the monomer rings and stitches them together one by one. Whereas molybdenum and tungsten catalysts are more active, ruthenium compounds can react selectively with a C=C bond in the presence of many other groups, such as ketones and esters. The tradeoff for such a tolerant catalyst is reduced reactivity toward low-strain rings, such as cyclopentene and cycloheptene, which are appealing substrates because they can be functionalized symmetrically to yield regioregular polymers.

Hejl *et al.* show that by careful tuning of catalyst and monomer concentrations, some of these rings can be coaxed open by Ru-based systems. In particular, they achieve >80% yields for polymerization of the unsubstituted 5- and 7-membered cyclic alkenes, and >60% yields for several ketone- and ester-substituted variants. The authors used density functional theory to calculate the strain in each monomer variant and found that the threshold for Ru-catalyzed ROMP is a minimal strain of 3.4 to 4.4 kcal/mol. — JSY

Macromolecules 10.1021/ma0501287 (2005).

HIGHLIGHTED IN *SCIENCE*'S SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

Science's stke www.stke.org

Signaling Behavior of Dopamine

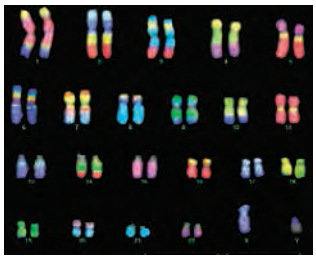
Dopamine is a neurotransmitter in circuits that convey reward and motivation, and abnormalities in dopamine signaling have been associated with mental illness. In particular, reduced function of the D2-type dopamine receptor (D2DR) is thought to contribute to schizophrenia, addiction, and mood disorders. Park *et al.* used a yeast two-hybrid screen to uncover prostate apoptosis response 4 (Par-4) as a binding partner for D2DR. In striatal neurons from mice that expressed a mutant form of Par-4 (in which the domain mediating the interaction with D2DR had been deleted), activation of signaling through cAMP was disrupted. Furthermore, behavioral tests of the mutant mice showed a depression-like phenotype, but no effects on measures of anxiety. Beaulieu *et al.* examined another signaling pathway emanating from D2DR, and they find that β-arrestin 2 is important in mediating the behavioral effects of dopamine. In wild-type mice, β-arrestin 2 was shown to associate with protein phosphatase 2a (PP2A) and the protein kinase Akt; this interaction increased after treatment with dopamine, which produced a decrease in Akt activity. In contrast, in mice deficient in β-arrestin 2, PP2A and Akt did not associate with D2DR, and dopamine did not affect Akt activity; this latter set of mice also showed decreases in dopamine-dependent behaviors. D2DRs are targets of antipsychotic drugs, so both studies provide hope that understanding the complexities of dopamine signaling may lead to the development of therapeutics that would be more effective and have fewer side effects. – NRG

Cell **122**, 275; 261 (2005).

Published by AAAS

NetWatch

edited by Mitch Leslie



EXHIBITS

Fun With Genetics

Dive into a human skin cell and zoom in on a loop of DNA. Quiz an expert about the genetics of diseases such as lupus. Those are two of the activities you can try at Understanding Genetics from the Tech Museum of Innovation in San Jose, California. Interactive exhibits let you explore topics such as eye color inheritance and whether the produce in your refrigerator could be genetically modified. The museum's on-call geneticist discusses issues in the news and answers questions from readers, such as whether a vegan diet reduces your tolerance for milk. Probably not, because the gene for lactase, the enzyme that breaks down milk sugar, naturally shuts off as most people age, regardless of diet. But not drinking milk for a while might eliminate bacteria that help digest it.

www.thetech.org/exhibits/online/ugenetics

FUN

Putting a Kick Into Physics

Unlike those aerial acrobats lofted by wires in martial arts movies, real kung fu experts depend on an implicit understanding of physics as they block and strike. For a lighthearted exploration of the connection between physics and martial arts, punch up the new exhibit Kung Fu Science from the Institute of Physics in London. Follow along as a physics student learns how to chop a board in half with her hand. Check out her calculations of how much energy her hand can apply to the wood, and then see if she can put the results into practice by breaking a stack of boards.

www.kungfuscience.org

TOOLS

Snooping for SNPs

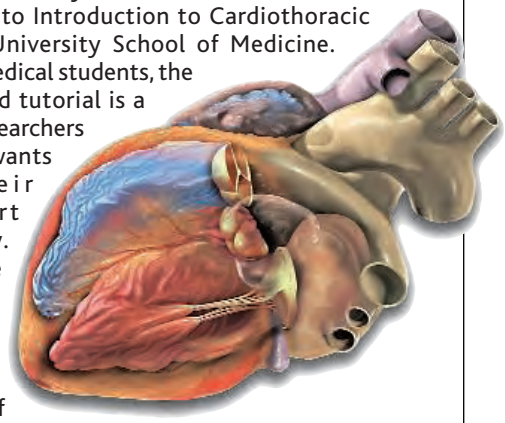
The genome's typos, single-nucleotide polymorphisms (SNPs), are one-letter changes in DNA that can signal susceptibility to diseases. WatCut from the University of Waterloo in Canada helps pinpoint SNPs in DNA samples. Users enter a SNP-containing DNA sequence, and WatCut identifies restriction enzymes that will chop the segment. The site can also hunt for silent mutations that allow a restriction enzyme to slice a sequence but that don't change the amino acids the sequence codes for.

watcut.uwaterloo.ca/watcut/watcut/template.php

IMAGES

Portrait of the Heart

Can't remember the location of the tricuspid valve? Need to know what an aortic aneurysm looks like on an echocardiogram? Click over to Introduction to Cardiothoracic Imaging from Yale University School of Medicine. Although aimed at medical students, the beautifully illustrated tutorial is a good resource for researchers or anyone else who wants to pump up their knowledge of heart and lung anatomy. Other sections use x-rays, echocardiogram footage, and other media to show how the structures change as a result of diseases such as emphysema and mitral stenosis, a narrowing of the opening between the left atrium and ventricle that can allow blood backflow. You'll also find a rundown of various imaging techniques.



info.med.yale.edu/intmed/cardio/imaging

RESOURCES

The Life Gelatinous

The world's longest animals don't have a mouth bristling with baleen or even a skeleton. Reaching 40 meters, the record-holders are siphonophores, relatives of jellyfish and corals. Get a peek at the marine predators, such as the deep-sea resident *Marrus orthocanna* (left), with this primer* from Yale grad student Casey Dunn. A siphonophore is a squishy commune, made of multiple units called zooids, each of which resembles an individual animal. Pages explain how one zooid gives rise to a siphonophore's elongated body. The site also showcases a new deep-sea species, which Dunn and co-workers recently reported in *Science*, that lures its prey with glowing tentacles (8 July, p. 263).

To snare more information about jellyfish and their kin, check out the Jellies Zone,† from curatorial associate David Wrobel of Harvard's Museum of Comparative Zoology. You can read up on groups of gelatinous creatures that live along the U.S. Pacific Coast, peruse a jellies FAQ, and browse galleries crammed with spectacular photos.

* www.siphonophores.org

† jellieszone.com

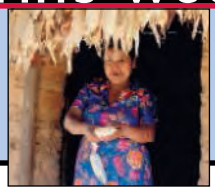
Send site suggestions to netwatch@aaas.org. Archive: www.sciencemag.org/netwatch

CREDITS (TOP TO BOTTOM): THE TECH MUSEUM OF INNOVATION; PATRICK J. LYNCH/YALE UNIVERSITY; CASEY DUNN

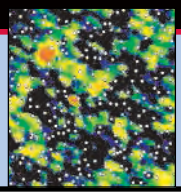
Published by AAAS

NEWS

This Week



PAGE 1000
Transgenes absent in Mexican corn



1001
Chaos in superconductors

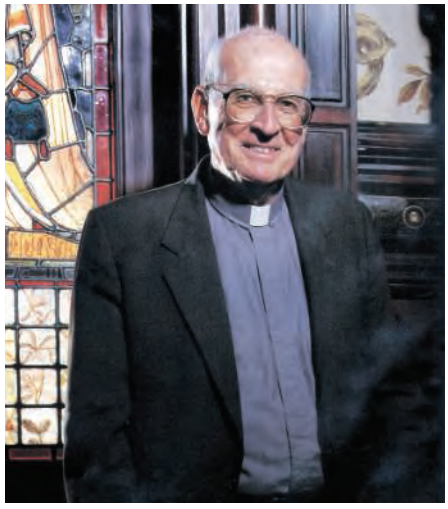
EVOLUTION

Vatican Astronomer Rebuts Cardinal's Attack on Darwinism

Is the Catholic Church rethinking its support for evolution? That's what Cardinal Christoph Schönborn, the archbishop of Vienna, suggested last month in *The New York Times* when he asserted that the church does not accept "neo-Darwinism." His 7 July opinion piece disturbed many scientists, especially those in the United States already worried about a resurgence of creationism and its "scientific" cousin, intelligent design.

Last week, with no utterance forthcoming from the new pope, the Vatican's chief astronomer George Coyne took it upon himself to rebut Schönborn. Writing in the 5 August edition of *The Tablet*, Britain's Catholic weekly, the Jesuit priest accused the cardinal of "darken[ing] the already murky waters" of the evolution debate. He also pointed out that the International Theological Commission under the presidency of Cardinal Ratzinger, now Pope Benedict XVI, issued a statement last year that saw no conflict between Darwin's ideas and the teachings of the Church.

In his *Times* piece "Finding Design in Nature," Schönborn last month dismissed as "vague and unimportant" the declaration of Pope John Paul II in 1996 that evolutionary theory is compatible with Catholic doctrine. "Evolution in the sense of common ancestry might be true," the cardinal wrote, "but evolution in the neo-Darwinian sense—an unguided, unplanned process of random variation and natural selection—is not."



Evolutionary face-off. Astronomer and priest George Coyne.

It didn't take scientists long to react. On 13 July, three figures prominent in defending the teaching of evolution in the United States sent a letter to the new pope urging him to reaffirm his predecessor's statement. In these "difficult and contentious times," wrote physicist Lawrence Krauss of Case Western Reserve University in Cleveland, Ohio, Francisco Ayala of the University of California, Irvine, and Brown University biologist Kenneth Miller, "the Catholic Church [must] not build a new divide … between scientific method and religious belief."

Biologist Peter Raven, head of the Missouri Botanical Garden and a member of the Pontifical Academy of Sciences, thinks scientists may have "overreacted" to Cardinal Schönborn's comments. In fact, Raven says, there is no evidence that the statement was cleared with the pope. It reflects "a pretty serious misunderstanding of what evolution is and what the church had done before," he adds. Raven doubts that Benedict, who was an honorary member of the Pontifical Academy before he succeeded John Paul II, is about to switch course. "The church has had the same view on evolution for about 75 years," he says. But Krauss is not so optimistic. "Based on what I've read about this pope," he says, "it's not at all clear" where he stands. Cardinal Schönborn's spokesperson Erich Laetenberger did not make the matter any clearer: "The cardinal only expresses what the church thinks about the issue," he told *Science*. ▶

AVIAN INFLUENZA

'Pandemic Vaccine' Appears to Protect Only at High Doses

This week, a U.S. health official trumpeted apparently good news: An ongoing trial suggests that a vaccine can protect humans from H5N1, the bird flu strain many worry may evolve into a pandemic. But some flu experts found the glass half-empty. The vaccine seems to works only at doses so high that the world's flu vaccine factories could not churn out enough to combat a pandemic, they say. Based on the preliminary data, the current U.S. stockpile of the vaccine, produced by Sanofi Pasteur, is enough for only 450,000 people—not the more than 2 million the administration hoped it would protect.

The findings, although an encouraging proof of principle, show the urgent need to develop ways to use vaccine more sparingly and to replace chicken eggs—the limiting step in current flu vaccine technology—with cell-based production systems, says Jeroen Medema, a vaccine scientist at Solvay, another flu vaccine producer. At the moment, he notes, "it's a vaccine for the happy few."

The new vaccine is based on an H5N1 strain isolated from a Vietnamese patient and genetically weakened to make it grow in eggs by researchers at St. Jude Children's Research Hospital in Memphis, Tennessee. Last weekend, Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases, which funded the trial, announced in newspaper interviews that initial data from 113 of the trial's 452 subjects show the vaccine eliciting protective antibodies. But to reach "levels that give you confidence," says Fauci, two doses of 90 micrograms of purified killed virus, or "antigen," had to be given 4 weeks apart.

The most common seasonal influenza vaccine is one shot of 45 micrograms of antigen—just 15 micrograms for each of the three circulating strains it targets. Because no one has immunity to H5N1, most researchers believed more than that might be needed in the new vaccine; plans for the U.S. stockpile were based on the assumption that two shots of 15 micrograms would work. "But 180—that really is a lot," says virologist Albert Osterhaus of Erasmus Medical Center in Rotterdam, the Netherlands.

Fauci says trials with dose-sparing strategies, including immune boosters called adjuvants, are next on the agenda. Many experts hope that with powerful adjuvants, a single dose of less than 2 micrograms of the vaccine might be enough, says Osterhaus.

–MARTIN ENSERINK

Published by AAAS

The academy's president, physicist Nicola Cabibbo of the University of Rome, has promised to look into the issue, says academy member and astronomer Vera Rubin of the Carnegie Institution in Washington, D.C. In an interview in the 18 July issue of the *National Catholic Reporter*, Cabibbo indicated that he endorses the views held by Pope John Paul II on evolution. Although some scientists think that "evolutionism" rules out God, Cabibbo declared, "this extension of Darwin's theory is not part of what has been discovered by science." Coyne makes reference to this debate in his recent essay, noting that "there appears to exist a nagging fear in the church" that the universe as defined by science "escapes God's dominion."

Meanwhile, defenders of evolution are still lamenting a comment last week by a vacationing President George W. Bush, in response to a reporter's question, suggesting that public schools should teach students about intelligent design (*Science*, 5 August, p. 861). Groups representing biologists, astronomers, and science teachers, among others, have shot off letters to the White House expressing their dismay. **–CONSTANCE HOLDEN**

With reporting by Eliot Marshall.

GENOMICS

Painstaking Approach Pays Off for Rice Sequencing Project

TOKYO—Finishing 3 years ahead of schedule and delighting agricultural researchers worldwide, a publicly funded international consortium has completed a highly accurate sequencing of the *japonica* rice genome. "It sets a gold standard" for plant sequences, says Hei Leung, a rice geneticist at the International Rice Research Institute (IRRI) in Los Baños, the Philippines.

The results, which appear in this week's issue of *Nature*, vindicate the International Rice Genome Sequencing Project's (IRGSP's) use of a time-consuming procedure in which the researchers created libraries of small bits of rice DNA and then sequenced them piece by piece. This map-based approach came under fire a few years ago after two teams not in the consortium published draft sequences of the rice genome based on a different technique (*Science*, 5 April 2002, p. 32). That approach, called whole-genome shotgun sequencing, busts the entire genome into different-sized bits, sequences them, and then uses supercomputers to put the data in order.

IRGSP researchers feared that their funding agencies would assume the job was done and pull the plug. But IRGSP leaders successfully argued that the drafts had too many gaps and errors to do justice to the world's most important cereal. "Supporting governments responded very positively, even increasing budgets to complete the map-based sequence," says Takuji Sasaki, director of Japan's Rice Genome Research Program, which led the IRGSP effort. Sasaki says he does not have a good estimate of the total project cost, but Japan spent roughly $100 million to sequence 55% of the genome.

Although the extra funding accelerated sequencing, Sasaki says the group finished ahead of its 2008 target date because of help from U.S.–based Monsanto, which had announced in 2000 that it would make its rice sequence data available to researchers and donate a library of bacterial artificial chromosomes (BACs), each with a fragment of rice DNA. "That was very important to us," Sasaki says. Syngenta, a Basel, Switzerland, agribusiness that had published a draft rice sequence in 2002 based upon the whole-genome shotgun approach, also contributed its data.

IRRI's Leung says a complete rice sequence will enable DNA microarray techniques to probe for thousands of single-nucleotide polymorphisms, minute genetic variations, across different rice varieties. Investigators hope to identify the combinations of genetic variations associated with complex traits such as drought tolerance. "The drafts would not allow us to use this technique," Leung says.

With the grain. Japan's Takuji Sasaki (*far left*) led an international effort to sequence the rice genome.

The benefits of a highly polished *japonica* genome sequence go beyond rice because other cereal crops, such as wheat and maize, tend to have the same genes in the same order. But the rice genome, at about 400 million bases, is much more compact than maize, which has about 2.3 billion bases. Joachim Messing, a molecular geneticist and director of Rutgers University's Waksman Institute of Microbiology in Piscataway, New Jersey, says a consortium working on the maize genome has used the IRGSP sequence to align BACs made for maize sequencing into a physical map of the maize genome much more efficiently and quickly than if they had started from scratch or used the two draft rice genomes. The accuracy and completeness of the IRGSP sequence is key, he says, because "if you have holes [in the sequence], then all comparisons with other genomes become tricky."

The other draft rice genome was of *indica*, a strain widely cultivated in China. Yu Jun of the Beijing Genomics Institute, a publicly funded Chinese institution that also published its draft genome in *Science*, says BGI has completed sequencing that first variety and is finishing a second *indica* variety. These two are the parental strains of a hybrid rice increasingly important in China. The data will help identify which genes are dominant in the first-generation crosses, which produce from 15% to 30% more grain than either parent. "We're looking for the secret of this hybrid vigor," he says. **–DENNIS NORMILE**

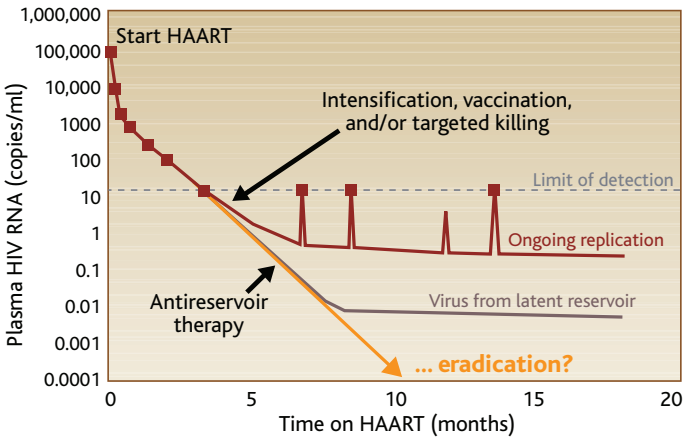
CREDITS (TOP TO BOTTOM): B.S.P.I./CORBIS; TAKUJI SASAKI/IRGSP

HIV/AIDS

Report of Novel Treatment Aimed at Latent HIV Raises the 'C Word'

A report in the 13 August issue of *The Lancet* is roiling the world of AIDS research. It describes an unusual treatment of four HIV-infected people, and the authors suggest that the strategy may point the way to a "cure of HIV in the future." The paper, co-written by a prominent collection of AIDS researchers, ventures into terrain that few have explored: eradicating the virus from all infected cells throughout the body. "The drumbeat we've been hearing for the last 5 years is this can't be done," says the study's leader, virologist David Margolis of the University of North Carolina, Chapel Hill. "The level of skepticism is very high. And rightly so. But the data we've gotten make me more hopeful."

In the AIDS field, few dare to use the "c word," and its use here has stirred criticism. Even "the words 'on the way to a cure' are just so inappropriate," says Robert Gallo, head of the Institute of Human Virology in Baltimore, Maryland, where Margolis once worked. "I think that was really a serious mistake." Researchers are wary in part because they have been burned before: Early enthusiasm in 1996 about the power of anti-HIV drug cocktails—called highly active antiretroviral therapy (HAART)—led David Ho of the Aaron Diamond AIDS Research Center (ADARC) in New York City to famously propose that eradication might take only 2 to 3 years of treatment. The concept lost currency when it became clear that HIV hides out in a latent state in cellular reservoirs from which it is very hard to dislodge.



Aiming low. Antiretroviral therapy knocks HIV below detectable levels (upper line); new therapy aims also to eradicate it from latent reservoirs.

In the new study, Margolis and co-workers recruited four patients who for 2 years had fewer than 50 copies of HIV per milliliter of blood, the detection limit of the standard assay. The researchers first drew huge amounts of blood from each patient through a process called leukopheresis and, using a highly sensitive test that purports to detect a single copy of HIV per milliliter, sifted through hundreds of millions of resting CD4+ white blood cells—the main harbor for latent virus—to assess the size of each patient's viral reservoir. They then gave the patients a drug they believed would force their resting cells to spit out new copies of HIV, which theoretically exposes the cells to immune attack or self-destruction. To help mop up bursts of the virus, the patients added a new drug, T-20, to their standard cocktails. After 4 months, the amount of infectious HIV in each patient's pool of latent cells declined an average of 75%, the investigators reported.

Researchers roundly praise Margolis for doing this difficult study, but reactions to the data and the postulated mechanism of action have been decidedly mixed. "It's an alternative approach that's worth pursuing," says Anthony Fauci, head of the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland. "But you have to be very careful about the hope you have for eradication with this. We went through the same thing a few years ago." Fauci, working with Tae-Wook Chun in his lab, found that they could reduce the size of the pool of latently infected cells in two patients, using interleukin-2 (IL-2) to "activate" the resting cells and thus flush the latent virus out of hiding. But as Fauci, Chun, and colleagues reported in the 28 October 1999 issue of *Nature*, when the patients stopped treatment, HIV quickly surfaced and refilled latent pools within weeks. Other groups have reported similar results.

The new study takes a more precise approach. Drugs such as IL-2 that nonspecifically activate white blood cells also create more CD4+ cells for the virus to infect. "You're always going to be chasing your tail" with such strategies, says Margolis. Along with Carine Van Lint of the University of Brussels and Eric Verdin of the Gladstone Institute of Virology and Immunology in San Francisco, Margolis has evidence that an enzyme called ▶

SOURCE: DAVID MARGOLIS/UNC

ScienceScope

Discovery Home Safe

To NASA's immense relief, the space shuttle Discovery arrived safely at Edwards Air Force Base in California 9 August after a 2-week mission to the international space station. But there is little time for celebration. Agency engineers are scrambling to solve the recurring problem of loose foam on the shuttle's external tank, which threatens the orbiter's delicate tiles. Although some observers inside and outside NASA speculate that the shuttle might never fly again, managers say they can fix the problem in time to meet a tight September launch window. Meanwhile, NASA Administrator Michael Griffin will begin a major lobbying effort in Washington, D.C., to win support for his postshuttle transportation plan, expected later this month (*Science*, 22 July, p. 540).

–ANDREW LAWLER

NIH Ethics Procedures Criticized

A review has found "vulnerabilities" in the way the National Institutes of Health (NIH) monitors scientists' outside consulting work.

Since February, NIH scientists have been under a ban on industry consulting while the agency puts new ethics procedures in place. The 72-page report by the Department of Health and Human Services' inspector general looked at outside activities approved for 174 senior employees between 2001 and 2003. Employees submitted "limited information" on their outside work, often forgetting forms or supervisors' signatures. But ethics officials fell short, too: 28% of activities were approved only after they began, for example.

–JOCELYN KAISER

Bottom-Dollar Sequencing

The U.S. National Human Genome Research Institute is betting on researchers to massively shrink the cost of sequencing large genomes. This week, it awarded more than $25 million to nine teams to develop technologies such as nanopores and molecular sensors that will speed the deciphering of DNA. The goal is a "$1000 genome" that will put sequencing machines into most labs and many medical clinics by the next decade. "We are ahead of schedule," says George Church of Harvard University, who has developed a sequencer that uses a microscope and other off-the-shelf equipment (*Science*, 5 August, p. 862).

–ELIZABETH PENNISI

Published by AAAS

histone deacetylase 1 (HDAC1) plays a crucial role in keeping CD4+ cells in a latent state. So Margolis gave his patients valproic acid, an HDAC1 inhibitor that's licensed to treat epileptic seizures. "It's more of a scalpel than the blunter instruments that we and others have used," agrees ADARC's Ho. "There are caveats about this study, but I'm certainly pleased to see their results."

Virologist Robert Siliciano of Johns Hopkins University in Baltimore stresses that even the 75% reduction of the latent pool has no clinical relevance. "Partial reductions of these cells sound good, but it's got to be complete to be useful," says Siliciano, whose lab specializes in HIV latency. He notes, too, that he and others disagree with Margolis about the mechanism of latency, questioning the value of HDAC1 inhibitors. "It's likely that there are several mechanisms," counters Margolis, who says this is just a small "proof of concept" study. Other researchers also caution that the assays used by Margolis and his collaborators are highly experimental, and it's unclear whether valproic acid or the intensification of HAART with T-20 were critical factors.

Roger Pomerantz, who also did eradication experiments in patients before he left academia to become president of the drug company Tibotec, says he's intrigued by the HDAC1 inhibitors, and he hopes the work spurs other clinical studies. Ho, who has taken heat for his earlier optimism, agrees. "It's OK to think about curing HIV," he says. "If we give up, there will never be a cure."

–JON COHEN

BIOTECHNOLOGY

Calming Fears, No Foreign Genes Found in Mexico's Maize

Mexico's transgenic maize scare appears to be over. This week in the online edition of the *Proceedings of the National Academy of Sciences* (*PNAS*), a team of Mexican and U.S. scientists report the results of a broad survey for foreign genes in native varieties of corn in southern Mexico. Four years ago, a report that such genes had been detected touched off an international furor. This time, scientists came up empty-handed: They detected no transgenes in seeds from hundreds of corn plants sampled in 2003 and 2004.

The negative results are good news for Mexican scientists and environmentalists, who worried that genes from genetically modified (GM) U.S. corn could contaminate the gene pool of Mexico's traditional varieties (landraces), conferring traits such as insect resistance that could skew their fitness. "The results will ease the concerns of many of us," says Universidad Nacional Autónoma de Mexico ecologist José Sarukhán, who was not part of the study. At the same time, the paper doesn't resolve lingering questions about whether foreign DNA was present in the first place.

Untainted. A new study finds no trace of foreign genes in traditional maize grown by indigenous farmers in Oaxaca, Mexico.

That issue exploded in late 2001 when biologists David Quist and Ignacio Chapela of the University of California, Berkeley, reported in *Nature* that they had detected genes from GM maize in four corn cobs collected in 2000 from the state of Oaxaca, part of the center of maize genetic diversity. Even more troubling, the genes were not always in their usual places; they appeared in random locations on chromosomes, suggesting that they could hop around and disrupt normal genes. Mexico had barred the planting of GM corn in 1998, so the reported transgenes suggested that farmers were illegally planting kernels from GM maize imported as food from the United States. Groups such as Greenpeace and the Mexican Congress subsequently called for a ban on imports of transgenic corn.

The controversy escalated when several molecular biologists questioned the study—particularly the claim that genes were jumping around. They noted that Chapela and Quist used the polymerase chain reaction, which is prone to false positives. In the face of this criticism, *Nature* asked the authors to submit more data using a different technique. The pair did, but the journal's editors were not convinced: They issued an unusual statement saying that the original paper should not have been published (*Science*, 12 April 2002, p. 236).

Meanwhile, government scientists had also detected GM genes in 5% or more of native corn samples from some fields. But when they tried to get the data published, reviewers were skeptical, says Exequiel Ezcurra, then president of the Instituto Nacional de Ecologia (INE) in Mexico City and one of the investigators. So the Mexican group, led by Sol Ortiz Garcia of INE, decided to start over. They also joined forces with ecologist Allison Snow of Ohio State University in Columbus, who has studied the risks of gene flow from other transgenic crops.

The scientists collected corn from 125 fields across a swath of Oaxaca in late 2003 and 2004 and sent pooled samples of more than 153,000 seeds from 870 cobs to two commercial labs in the United States. The tests found no traces of foreign DNA in the Oaxaca samples, nor in more limited samples of other regions. If the transgenes are present, the levels are below 0.005%, the limit for detection. The results were a surprise. "We were expecting to find transgenes," says Snow.

So what happened to the foreign DNA apparently detected in 2000 corn? The authors suggest that an education campaign may have deterred Oaxaca farmers from planting more GM kernels, and offspring of any transgene-tainted plants may not have done well in Oaxaca's mountains. It's also possible the foreign genes were never present. Ezcurra, now at the San Diego Natural History Museum, believes they were: "I don't think so many labs could have found positives without something going on there." But the Mexican scientists didn't save their 2000 maize samples, he says, so the question may never be settled. Chapela offers another explanation—the sampling and testing methods used in the new study may have missed extremely low levels of transgenes. Otherwise, he says, "it's very hard to make both [papers] compatible."

Assuming transgenes were present but disappeared, that good news is no reason for Mexico to relax, say several scientists. The country could soon allow GM field trials of maize, and strict biosafety rules will be essential, Sarukhán says. Moreover, a "massive flow of maize" continues from the United States, and chances are that GM corn is growing elsewhere in Mexico, says Greenpeace scientific adviser Doreen Stabinsky. Last year, the U.S. and Mexican governments rejected a suggestion from a panel of North American biodiversity experts that Mexico require that U.S. corn be ground up before it is imported. "That's the kind of complacency you don't want this *PNAS* paper to generate," Stabinsky says.

–JOCELYN KAISER

CREDIT: *PNAS*

PHYSICS

Physicists Get the Dope on Disorder in High-Temperature Superconductors

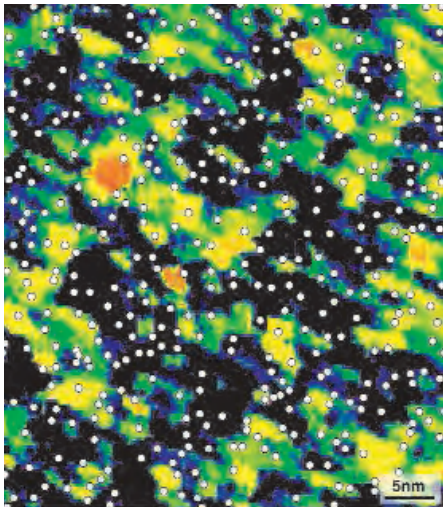
Within high-temperature superconductors—layer-cake materials that carry current without resistance at temperatures as high as 130 kelvin—chaos reigns. To the vexation of physicists, key properties of the material vary randomly from place to place inside all but the most pristine samples. Some argue that the atomic-scale variations are a necessary ingredient of the superconductivity; others think they have nothing to do with it. Now, measurements reported on page 1048 reveal the source of much of the disorder: It's caused by oxygen "dopant" atoms strewn throughout the layered crystals—the same atoms that supply electric charges for the superconducting currents.

By linking the disorder to the oxygen atoms, the results strike a blow against theories in which the variations arise spontaneously from interactions of the charges and then help the charges flow without resistance, says Elbio Dagotto, a theorist at the University of Tennessee, Knoxville. "This result is telling me that some of the more exotic scenarios may not materialize," he says. But John Tranquada, an experimenter at Brookhaven National Laboratory in Upton, New York, says the meaning of the new observations isn't obvious. "Does this tell us something deep about the mechanism of high-temperature superconductivity?" he says. "I don't see a clear message on that."

Kyle McElroy and J.C. Séamus Davis of Cornell University in Ithaca, New York, and colleagues made the measurements by taking a close-up look at the superconductor bismuth strontium calcium copper oxide, or BSCCO (pronounced "bisco"). BSCCO consists of planes of copper and oxygen atoms arranged in checkerboardlike squares, interspersed with layers containing other elements. The dopant oxygen atoms nestle between the planes, soak up some of the negatively charged electrons in them, and leave behind positively charged "holes" that pair to carry current without resistance along the copper-and-oxygen planes. However, physicists had never been able to pinpoint the location of the dopant atoms to see precisely how they affect the flow of electricity.

To do that, the Cornell team scanned BSCCO samples with a scanning tunneling microscope (STM), a probe just a few atoms wide from which current flows into the surface of a material. When the researchers applied large negative voltages from tip to surface, they noticed that the current shot up in specific locations. Less than a nanometer wide, the spots were more common in samples doped with more oxygen, indicating that each marked the location of an oxygen atom.

At smaller voltages, the researchers also traced out the so-called superconducting gap—a range of voltage at which the current dwindles—and the adjoining coherence peaks, ranges in which the current shoots up again. The size of the peaks and the width of the "valleys" between them varied in ways that were correlated with the positions of the oxygen atoms, Davis says, suggesting that the atoms cause the seemingly chaotic fluctuations that have puzzled physicists.



Spotted. In a high-temperature superconductor, oxygen dopants (white spots) lie in regions where the "superconducting gap" grows wider (dark regions), new measurements show.

The peaks shrank near each oxygen dopant, possible evidence that each oxygen atom damages the superconductivity in its neighborhood, Davis says. If so, it may be possible to achieve superconductivity at higher temperatures by doping the material in a more clever way, he says—perhaps by eliminating the oxygen dopants and using a strong electric field to pull electrons out of the copper-and-oxygen planes.

But others are cautious about the speculation that the oxygen dopants damage the superconductivity around them. "If it's true, then I guess future work will show it," says Øystein Fischer, an experimenter at the University of Geneva, Switzerland, who also does STM experiments. Still, by tying the electronic disorder to the dopants, the results make an important contribution to the study of high-temperature superconductors, Fischer says. Things are still a mess within them, but now researchers know why. **–ADRIAN CHO**

CREDIT: K. MCELROY *ET AL.*, *SCIENCE*

ScienceScope

Uzbeks to Scientists: Hit the Silk Road

An international conference devoted to the disappearing Aral Sea has fallen victim to a Soviet-style freeze. In an 11th hour snub to the European Union, the Ministry of Foreign Affairs (MFA) of Uzbekistan last week refused to grant visas to participants in the Aral Sea Basin Water and Food Conference, scheduled for early next month in Tashkent. Relations between Uzbekistan and the West have soured since May, when the government suppressed an uprising in the eastern city of Andijan and blocked an independent inquiry.

The scuttled conference was to cover issues such as managing scarce water supplies and growing salt-tolerant crops in the exposed lakebed of the Aral Sea, which has shrunk by 75% since the 1960s (*Science*, 18 February, p. 1032). The topics are "of uttermost importance," says John Lamers, a senior researcher at the University of Bonn in Germany, who was to chair a panel on agriculture. MFA said the conference had not been approved by the Uzbek Cabinet; the rebuff prompted the sponsor, INTAS, a Brussels-based fund that supports science cooperation with the former Soviet Union, to pull the plug after being unable to arrange an alternative venue.

–RICHARD STONE

Report Seeks Delay on Waste

The Department of Energy (DOE) should wait before sealing radioactive waste tanks at the Savannah River Site in South Carolina, a National Academies panel said in an interim report last week. For years, activists tried in court to force DOE to rid the weapons-building site's 51 underground tanks of all nuclear waste. DOE has argued that permanently sealing some waste in place with grout can be environmentally sound and cost-effective, and last year lawmakers gave it the authority to do just that.

But the congressionally mandated report says that postponing permanent closure on hard-to-clean tanks would have "no effect on near or long-term risk" and could give researchers time to improve cleanup methods within a decade. DOE, which calls it unwise to postpone tank closures to wait for new technologies, was unable to provide the panel with many of the requested documents, citing internal reviews. But panel members hope to have some of that information in time for their final report due in January, which will examine cleanups in Washington state and Idaho as well. **–ELI KINTISCH**

Published by AAAS

With no vaccine in sight, AIDS researchers are testing a range of surprising biomedical interventions

Prevention Cocktails: Combining Tools To Stop HIV's Spread

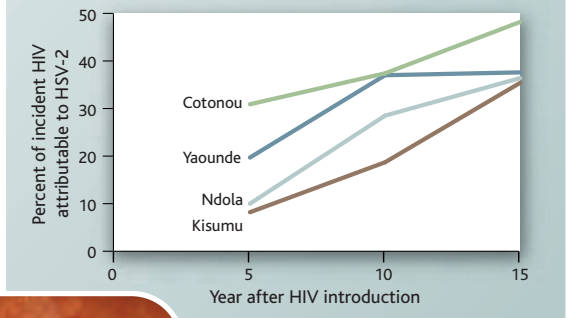
If AIDS researchers reported that a vaccine protected 65% of the participants in an efficacy trial, the news would be trumpeted across the globe. Two weeks ago at an AIDS meeting in Brazil, a study revealed that male circumcision produced that level of protection in South Africa. Many major media did not even mention this advance.

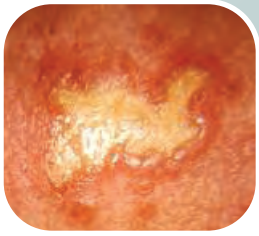
True, male circumcision as an HIV prevention strategy pales in comparison to a vaccine, a few shots of which theoretically could train the immune system of both genders to ward off HIV for decades. But the search for a safe and effective vaccine has stumbled repeatedly, and fundamental questions remain about whether a vaccine is even feasible, much less how it would work. These frustrations have prodded researchers to explore other, decidedly more mundane, alternatives like circumcision.

Nearly a dozen potential preventives are now under study that have a refreshing simplicity to them. They include drugs already on the market, existing devices such as the female diaphragm, and such basic concepts as improving genital hygiene. The hope is that these could work together with condoms and behavior change to help communities slow AIDS epidemics. "We all know that abstinence and couples being mutually faithful would be great if they were applicable to everybody's lives, but they're not," says Helene Gayle, who directs the HIV, TB, and Reproductive Health program for the Bill and Melinda Gates Foundation. "These more short-term endeavors are giving people hope. We know that's its going to take at least decade to get to a vaccine." Adds psychologist Thomas Coates, who does prevention research at the University of California, Los Angeles (UCLA): "It's a new era of prevention."

Each of these interventions, circumcision included, has serious limitations. They also could do more harm than good if they lull people into taking more sexual risks. That's just one of several vexing ethical dilemmas that prevention researchers are facing. But Gayle, who has helped steer the Gates Foundation's funding of many of these projects, says the promise is undeniable. "People are energized in ways that they weren't before," says Gayle. "People have gotten jazzed."

Infection with HSV-2 paves the way for HIV





Co-factor. Infection with herpes simplex virus 2 (*left*) in four African cities is estimated to have accounted for more than 30% of new HIV infections.

Beyond observation

In addition to the vaccine field's travails, the impetus for many of the new interventions being tested comes from observational studies that have highlighted the co-factors most responsible for HIV transmission. "There are interesting scientific data that support development of very tightly reasoned biological hypotheses that are not just relying on a vaccine," says Kenneth Mayer, director of the Brown University AIDS Program in Providence, Rhode Island, who does prevention studies in several countries. Roughly 5 years ago, two large observational studies began to yield several overlapping insights.

One, the so-called Study Group on the Heterogeneity of HIV Epidemics in African Cities, looked at 8000 men and women from four locales, two of which had much higher HIV prevalence than the others. Anne Buvé, an epidemiologist at the Institute of Tropical Medicine in Antwerp, Belgium, and her colleagues found that circumcision and preexisting infection with herpes simplex virus-2 (HSV-2), which causes genital ulcers, seemed to account for much of the difference in prevalence. The second study, led by Ronald Gray of Johns Hopkins University in Baltimore, Maryland, and Maria Wawer of Columbia University in New York City, followed 15,000 adults in the Rakai District in Uganda. The researchers found that in "discordant" couples in which only the woman was infected with HIV, if the male partner was circumcised, which occurred in 50 cases, she never transmitted the virus; nearly 17% of the uncircumcised men did acquire the virus from their infected partners. In these same initially discordant couples, people with higher HIV levels—or viral loads—more readily spread their infection. And the researchers later found that HSV-2 infection strongly increased the likelihood of transmission.

Both the four-city and Rakai studies have become landmarks in the field, and clinical trials are now building on those observations. A lead investigator in the four-city study, Bertran Auvert of the Uni-

CREDITS (TOP TO BOTTOM): MALCOLM LINTON; ESTHER FREEMAN/ LONDON SCHOOL OF HYGIENE AND TROPICAL MEDICINE, INSET: JOHN F. WILSON, MD/PHOTO RESEARCHERS INC.

Published by AAAS

Testing patience. A social worker (red sweater) in Kolkata, India, takes a group of people to the city's busy HIV testing clinic at the School of Tropical Medicine.

versity of Versailles in Saint-Quentin, France, headed the South African trial that found 65% protection from circumcision. Gray and Wawer are currently running a similar circumcision study in HIV-uninfected men in Rakai, as well as a second trial that asks whether circumcising HIV-*infected* men in discordant couples might reduce transmission. (Yet another circumcision study underway in Kisumu, Kenya, run by Robert Bailey from the University of Illinois at Chicago School of Public Health, is also evaluating circumcision of HIV-uninfected men.)

A model based on data from the four-city study underscores circumcision's potential to alter AIDS epidemics. As Kate Orroth from the London School of Hygiene and Tropical Medicine reported last month at an Amsterdam conference on sexually transmitted diseases (STDs), her preliminary data suggest that if circumcision rates jumped from 10% to 100% in the Zambian city of Ndola, the prevalence of HIV in adults would drop from 27% to 7% in little more than a decade—and that's assuming circumcision offers only 50% protection.

Following up on the HSV-2 lead, epidemiologist Connie Celum from the University of Washington in Seattle is heading two multisite, international trials of daily acyclovir, which is licensed to treat herpes infections, to see whether suppressing that virus can reduce the incidence of HIV transmission. "These trials have a reasonable chance of providing some data that will reshape our focus on HIV and sexually transmitted diseases," says Celum. One trial will include some 3000 HIV-uninfected people. The other, building on evidence that HSV-2 reactivation helps HIV copy itself—and thus makes a person more infectious—is recruiting 3600 couples who are discordant for the AIDS virus.

Acyclovir is ideal for this type of study because it "has virtually no toxicity except in really high doses," says Celum, and there's little danger that daily doses will lead to the emergence of drug-resistant strains. For HSV-2 to become resistant to acyclovir, it must mutate a key enzyme used by the virus, which reduces its "fitness," Celum explains. She knows of only two cases in which people transmitted such resistant strains.

If acyclovir treatment of HSV-2 works as an HIV prevention strategy, it too could greatly affect AIDS epidemics. HSV-2 infects from 22% of adults in the United States to a staggering 70% of women in southern Africa. And that's in HIV-*un*infected people; more than 80% of HIV-infected adults are co-infected with HSV-2. Again, models offer provocative predictions. At the Amsterdam STD meeting, Esther Freeman, a grad student who works with Orroth and Richard Hayes at the London School of Hygiene, used the four-city data to show that 15 years after HIV was introduced to those locales, HSV-2 accounted for more than one-third of the new infections with the AIDS virus (see graph, p. 1002). "It's a huge effect," Freeman says.

Direct hit

Other prevention trials underway use anti-HIV drugs to attack the virus directly.

Antiretrovirals lower viral loads, and given the Rakai data showing that people with less virus are less infectious, this strongly suggests that anti-HIV drugs might work as both a treatment and prevention tool—but that remains to be proved. "Knowing whether they have some benefit in prevention is a really important question," says Brown University's Mayer.

To specifically address this question, the HIV Prevention Trials Network (HPTN), sponsored by the National Insti-

Hedged Bet: An Unusual AIDS Vaccine Trial

Even the AIDS vaccine world has jumped on the simplicity bandwagon. To many AIDS vaccine researchers, the key obstacle is that no one has yet found a vaccine that can trigger effective antibodies against the surface protein of the virus. So Merck has constructed a vaccine that abandons antibodies altogether, and the company is testing it in a fast-tracked study to determine whether it's worth pursuing the approach.

Although antibodies prevent cells from becoming infected, the Merck vaccine attempts to train the cell-mediated arm of the immune system, which eliminates cells that HIV has infected. The vaccine uses adenovirus to carry three HIV genes, but, in a marked difference from almost every other vaccine under development, not the gene for the surface protein.

Working with the National Institute of Allergy and Infectious Diseases (NIAID) in Bethesda, Maryland, Merck has launched a study in 3000 people at high risk of becoming infected. This unusual study is essentially a hedge bet: it will not have the statistical power of the typical Phase 3 efficacy trial that leads to licensure, so researchers are calling it a Phase 2b. "What do you do if you want to know if something works, and the only way to do it is humans, and you don't have enough confidence to do a Phase 3 study?" asks Peggy Johnston, who heads NIAID's AIDS vaccine program. "You do an overpowered Phase 2."

The trial aims to answer two discrete questions. First, most people have been infected with the adenovirus subtype (called Ad5) that Merck uses, and their antibodies against this "vector" could prevent it from producing the HIV proteins needed to stimulate a robust immune response. So half the people recruited for the international study, called Step, will have low levels of antibodies to Ad5. If the vaccine works, researchers then can evaluate whether the Ad5 antibody levels have any impact. Secondly, if it produces robust cell-mediated immunity, they'll know once and for all whether that response by itself can protect against HIV. "The Step trial is a good name for it," says Johnston. "I see it as a step forward. But it's not the final step." **–J.C.**

Fight on. The San Francisco Department of Public Health uses this ad to recruit for the Step study.

CREDIT: COURTESY OF HIV VACCINE TRIALS NETWORK

Published by AAAS

tute of Allergy and Infectious Diseases (NIAID) in Bethesda, Maryland, recently launched an ambitious antiretroviral treatment study led by Myron Cohen of the University of North Carolina (UNC) in Chapel Hill; it ultimately hopes to enroll 1750 discordant couples on four continents. Columbia's Wawer is also examining the role of antiretrovirals as a prevention strategy with a new, multiyear observational study in Rakai. Wawer essentially is taking advantage of the fact that the U.S. government is providing treatment to many HIV-infected Ugandans as part of the President's Emergency Plan for HIV/AIDS Relief.

The most advanced trials to test whether antiretrovirals can prevent infection involve giving a drug called tenofovir to uninfected people. Several monkey studies have shown that tenofovir—which cripples an enzyme that HIV needs to copy itself and has been on the market since 2001—works remarkably well at what's called pre-exposure prophylaxis (PrEP). In these experiments, researchers give animals the drug and then attempt to infect them with SIV, the simian relative of HIV. Monkeys that receive the drug up to 2 days before this SIV "challenge" have dodged the infection. Although enthusiasm dampened for this approach when a recent study from the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, showed that tenofovir-treated monkeys eventually did become infected after repeated challenges, many researchers suspect PrEP will work to some degree in humans. "In the absence of a vaccine, it could be a very effective tool against HIV," says UCLA's Coates.

Novel Prevention Studies

Intervention	HIV status	Product	Sponsors	Locations	Stage	Projected end
Pre-exposure prophylaxis	-	Tenofovir	CDC, Gates Foundation	Malawi, Ghana, Thailand, U.S., Peru	Safety–efficacy	2006–08
Circumcision	-	n/a	NIAID	Kenya, Uganda	Efficacy	2007
Circumcision	+	n/a	Gates Foundation	Uganda	Efficacy	2008
Treat HSV–2	-	Acyclovir	NIAID	Zambia, Zimbabwe, S. Africa, U.S., Peru	Efficacy	2007
Treat HSV–2 in discordant couple	+/-	Acyclovir	Gates Foundation	Kenya, Tanzania, Uganda, Zambia, Botswana, S. Africa	Efficacy	2008
Treat HIV in discordant couple	+/-	ARVs	NIAID	Malawi, India, Zimbabwe, Brazil, Thailand, U.S.	Pilot	2006
Diaphragm	-	n/a	Gates Foundation	S. Africa, Zimbabwe	Efficacy	2007
Topical penile microbicide	-	Ethanol based	NIAID	Kenya	Phase 2	Start in fall
Rectal microbicide	+/-	Biosyn UC-781 RT inhibitor	NIH	U.S.	n/a	Pending
Acute infection	+	PCR on pooled blood samples	NIH	Malawi, North Carolina	n/a	Pending
Vaginal microbicide	-	Tenofovir	NIAID	U.S., India	Safety	Start in fall
Treat depression	+/-	Bupropion	NIH	San Francisco	Phase 2	2006

Seven clinical trials, funded separately by the Gates Foundation and CDC, are now evaluating the safety and efficacy of tenofovir PrEP. Two other tenofovir PrEP studies ended prematurely after activists raised ethical concerns—which had more to do with trial designs than the specific intervention—and a third closed up shop for technical reasons (*Science*, 18 March, p. 1708). In an unusual twist, tenofovir's maker—Gilead Sciences of Foster City, California—says it has no interest in pursuing PrEP because of fears that uninfected people who take tenofovir and still become infected might sue the company.

Tenofovir appears to be safer than most antiretrovirals on the market, and if it works, it offers clear advantages over some other prevention strategies. "The idea of doing circumcision on a mass scale is kind of daunting," says Coates. "Providing pills is a lot simpler." Tenofovir PrEP might also work equally well in both sexes and isn't limited to people who already have another infection, like HSV-2.

Researchers have also begun to incorporate tenofovir and other antiretrovirals into microbicides, gels or creams that are put into the vagina—or, in one new study, the rectum. The five efficacy trials now underway with vaginal microbicides all rely on nonspecific formulations such as buffering agents and detergents; as a result, many researchers question whether any will have much success. These non-specific compounds must also be used about one hour before intercourse. "Maybe the deck is stacked against them," says Zeda Rosenberg, a virologist who heads the International Partnership for Microbicides in Silver Spring, Maryland.

In contrast, tenofovir and some other anti-HIV drugs—including one being developed by Rosenberg's nonprofit—remain active longer and may only need to be used once a day. And ideally, she says, microbicides will take a page from the treatment world and use a cocktail of anti-HIV drugs to attack the incoming virus from many angles at once.

Early containment

Very early detection of HIV infection may also offer an opportunity to prevent transmission when the risk is highest—which typically occurs before people even know they are infected.

The Rakai study and several since then have reported that people have the highest viral loads, and are most infectious, right after they become infected—and before infections show up in antibody tests. "You're never going to be able to deal with the epidemic until you deal with those acutely infected people," explains UNC's Cohen.

He and Christopher Pilcher have pioneered a strategy to better identify acutely infected people. They have used the polymerase chain reaction (PCR) to detect HIV in blood that has been pooled from thousands of people visiting STD clinics and the like. If they detect the virus, they break the pool into smaller and smaller pools for retesting, eventually identifying the individual patients who harbor the virus. As the researchers explained in the 5 May *New England Journal of Medicine,* they used this technique, which cost less than $4 per blood donor, in North Carolina to identify 23 acutely infected people. They and 48 of their sexual partners were notified and given counseling about how to

SOURCE: HPTN, GATES FOUNDATION, NIH

lower their risks. Twenty of the acutely infected people also opted to start treatment, likely reducing their viral loads.

Low tech

In this new era of prevention, even the commonly used diaphragm and other simple approaches are playing a role.

"It took me over 10 years to get this funded," says Nancy Padian, an epidemiologist at the University of California, San Francisco (UCSF), describing her study of the diaphragm and a lubricant as an HIV prevention device in Zimbabwe and South Africa. "People are interested in a new microbicide, a new vaccine. But the diaphragm? 'No, no, no,' " says Padian, who finally received funding from Gates.

As Padian explains, the diaphragm should prevent HIV from reaching the cervix and endocervix, where most female infections occur. If it works, she says, the diaphragm will have a distinct advantage as it will enable a woman to protect herself without having to negotiate with a partner, as often happens with condoms.

In males, basic hygiene of the penis may prevent transmission. King Holmes from the University of Washington in Seattle, working with Elizabeth Bukusi at the Kenya Medical Research Institute, is studying whether wiping the penis with an ethanol-based gel—similar to the commercially available Purell—can thwart transmission of HIV, HSV-2, and other sexually transmitted microbes. "There was a long history of men using topical prophylactics, but with advent of antimicrobials around World War II, these basically stopped," says Holmes.

One of the most provocative, low tech prevention studies focuses on the master organ that makes people vulnerable to HIV: the brain. Grant Colfax at the San Francisco Department of Public Health studies the link between methamphetamine use in gay men and HIV transmission. Meth users have decreased dopamine levels in the brain, which can lead to depression. Because studies have shown strong links between depression in gay men and sexual risk-taking, Colfax explains, he plans to launch a study this fall that will assess the impact of an antidepressant, bupropion (trade name Wellbutrin), which acts by indirectly increasing dopamine levels.

Real world

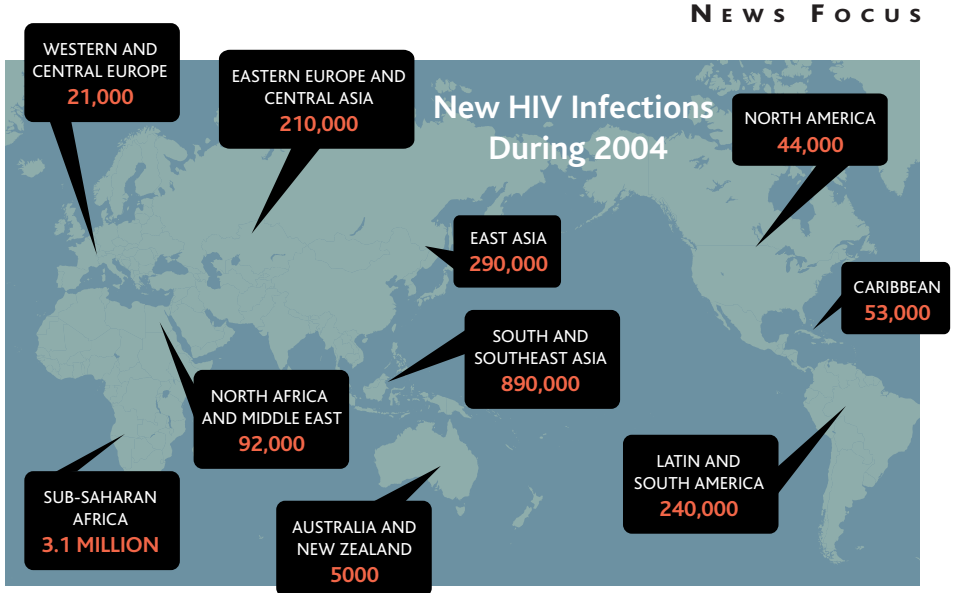
Researchers concede that it's difficult to envision how these myriad prevention interventions will play out in the real world. After all, the benefits of condoms have been widely known for years. In addition, clinical trials often fail to reflect how a drug is actually used. The tenofovir PrEP and acyclovir studies evaluate daily dosing, for instance, but if they work, people might take the drugs intermittently. More troubling still, investigators worry that the benefits of most prevention interventions could be undermined by what psychologists call behavioral disinhibition. Specifically, if an intervention—whether it's proven to work or still in trials—leads people to think they are protected and thus can safely have more sexual partners or unsafe sex, the risk of becoming infected could increase.

Another untidy dilemma is that success comes at a cost. If, say, tenofovir PrEP works, then ethics demand that everyone in any prevention study be offered the drug. "It could have a major impact on vaccine studies," says Peggy Johnston, who heads AIDS vaccine research for NIAID. "Each prevention tool that's added makes it harder for the next one to prove efficacious. But that's not necessarily a bad thing. It means were getting better prevention tools."

And in prevention, as in treatment, combining interventions appears to be the name of the game. "We need to really look at how we put together a combination of options that fit people's lifestyles," says the Gates Foundation's Gayle. Ideally, prevention campaigns also will promote these options to the high-risk groups most likely to become infected and spread the virus (a concept that has far less importance in the many sub-Saharan locales that have double-digit prevalence, which makes all sexually active adults vulnerable).

UCSF's Padian contends that a cocktail of the various prevention interventions now in trials could be "extraordinarily successful," but she notes that other than circumcision, each one requires that people repeatedly take actions to protect themselves. Which means that a safe and effective AIDS vaccine will remain an urgent need. But in the meantime, if more of these unflashy biomedical alternatives prove their worth, they could powerfully slow HIV, which now infects another 14,000 people—half of them between 15 and 24 years old—each day.

–JON COHEN



World of trouble. HIV infects nearly 5 million people each year.

Changing behavior? Limited success with prevention campaigns like this one being designed in Kunming, China, have fueled the search for new biomedical interventions.

CREDITS (TOP TO BOTTOM): SOURCE: UNAIDS; MALCOLM LINTON

Published by AAAS

Beloved Arctic Station Braces for Its Own Climate Change

Researchers of all stripes have been monitoring the impact of climate change at Alaska's Toolik Lake for decades. They now face new challenges

TOOLIK LAKE, ALASKA—Lying on his stomach, ripping moss off a two-by-three-meter patch of tundra, Tom Crumrine learned this summer what it means to study climate change in the Arctic. The high school teacher from Concord, New Hampshire, spent 3 weeks of his break on a fellowship at the Toolik field station on Alaska's North Slope, contributing to an experiment on how changes in vegetation caused by global warming will affect the Arctic landscape.

The answers won't be clear for years to come. But delayed gratification is the norm at Toolik Lake, where for 3 decades scientists have journeyed 300 kilometers above the Arctic Circle to assess the state of flora and fauna as Earth warms. "It's always been a place where you could come and really focus on your work," says Gaius Shaver, an ecologist at the Marine Biological Laboratory in Woods Hole, Massachusetts.

During the short research season, hundreds of ecologists, geomorphologists, plant physiologists, and biochemists traipse out to field sites, some 250 kilometers away, to measure and manipulate nutrients, species composition, temperature, water flow, and other factors. Their observations will be plugged into the still emerging picture of the impact of global climate change on Arctic ecosystems, which are coping with air temperatures that have risen by an average of 3°C in Alaska over the past 40 years.

The research station itself is also changing with the times. Toolik's leaders want to keep the lab open year-round, instead of just 4 months during the late spring and summer. They also hope to improve lab conditions, extend their studies into additional territory, and increase outreach to undergraduates as well as teachers like Crumrine and their students. "The Arctic is such a data-poor place," says Tom Pyle, director of the Arctic research section at the National Science Foundation (NSF), which provides the station with 90% of its $10.5 million in annual support. But its record, he says, makes Toolik "the crown jewel."

However, scientists fear the gem could lose some of its luster. Toolik sits just off the Dalton Highway, a road built and maintained for the oil wells 200 kilometers to the north at Prudhoe Bay, and there are plans for a pipeline to carry natural gas from those same fields. Other proposed developments could also encroach on ongoing or planned experiments. And scientists are worried about NSF's ability to continue to support their work. Flat budgets expected for the agency's 26 ecological research stations, of which Toolik is one, for ecosystems studies and for Arctic research jeopardize existing activities and leave little room for growth.

For science's sake. In June, University of Alaska researcher Syndonia Bret-Harte and teacher Tom Crumrine spent their days pulling moss to simulate potential global climate change effects.

"Our concern is Toolik has a unique dataset that we feel is very important," says John Hobbie, director of the ecosystems center at the Marine Biological Laboratory in Woods Hole, Massachusetts. If funders pull the plug, even temporarily, he warns, decades of work might be lost forever. "We can't reconstruct that data set."

Meager beginnings

It was 30 years ago that Hobbie first recognized Toolik Lake's potential as a research site. He followed the bulldozers as they built the "haul" road northward to bring pipes, gravel, and other construction goods to the oil fields. A 25-meter deep lake perfect for comparing lake and pond nutrient cycles, Toolik allowed Hobbie to further his NSF-funded research on aquatic ecosystems, a project he's continued ever since. A year later, Shaver showed up to study revegetation of the roadside, helping to establish a terrestrial complement to Hobbie's work.

With the Brooks Range as a backdrop, researchers at Toolik make use of continuous sunlight during the summer to work alongside caribou, bear, and moose on one of the world's most carbon-rich soils. Arctic tundra and boreal ecosystems take up one-sixth the world's land, but possess one-third the world's terrestrial carbon. Within hiking distance are tundra habitats ranging from wet soils covered with squishy moss to dry heath landscapes. The age of the land varies from 12,000 to 200,000 years, making the area a good place to understand how soils contribute to tundra ecology.

In 1987, the field station became an NSF Long-Term Ecological Research (LTER) site, directed by Hobbie. NSF core funding, now $820,000 a year, provides for technicians and equipment to help keep long-term studies going with or without income from individual investigator grants. The site put down even deeper roots in 1999, when NSF signed a cooperative agreement with the University of Alaska to run the station. A new 5-year agreement starting next year will provide about $1.5 million per year.

For years the lab grew slowly, retaining its rustic atmosphere. Two trailers served as kitchen, dining area, office, and lab. Scientists slept in backpacking tents. Today, there are four custom-built doublewides for labs, a dining hall open 24-7, a trailer for showers, and even one for laundry—32 buildings in all. There's a helicopter, and a fiber optic line put in for the Alyeska Pipeline Company of Anchorage provides Internet access. "They have put a lot of money into getting a really high-tech infrastructure," says Jeff Dudycha, an evolutionary biologist at William Paterson University in Wayne, New Jersey, who visited Toolik this summer as

CREDIT: E. PENNISI/*SCIENCE*

Published by AAAS

part of a field trip arranged for the Evolution 2005 meeting in Fairbanks. Adds Breck Bowden, an aquatic ecosystems ecologist from the University of Vermont, Burlington, "We have lab facilities that rival what I have [at my university]."

The terrestrial research at Toolik consists in large part of parallel experiments in different types of tundra. Plots are fertilized with either nitrogen, phosphorous, or a mixture of both. Some are housed in plastic greenhouses that boost ambient temperatures by an average of 3.5°C and simulate global warming, while others are shrouded in layers of greenhouse shade cloth that block half the incoming light. Fine chicken wire helps exclude small herbivores, and taller fences keep out moose and caribou.

The message from all these experiments is clear: The availability of nutrients is the driving force for the ecosystem. With the light cut by 50%, "there isn't much impact," says Shaver—at least not right away. The same is true for temperature. On acidic, 60,000-year-old tundra that has been fertilized, stands of dwarf birch eventually dominate, replacing sedges. The birch trees use added nutrients to grow taller and bushier, blocking ever more light from competitors. The shrub cover also insulates the ground, altering the season during which decomposition—and the release of nutrients—can occur.

In 2004, Shaver and his colleagues reported the surprising result that supplementary nitrogen and phosphorous fertilizer had increased plant production but resulted in a net loss of nitrogen and carbon from deeper layers of acidic soils. When exposed to fertilizer, soil microbes boosted their breakdown of the organic matter in those layers. Moreover, fertilizer increased the rate at which organic nitrogen was converted to an inorganic form. He also learned that productivity in greenhouse plots can change: It took 9 years to see an increase in productivity in the unfertilized greenhouse plots, as the increased temperature slowly boosted microbial activity and the release of nutrients locked up in the soil. "The message I get is to be careful about jumping to conclusions," says Pyle.

Research on streams and their role in nutrient flow is also yielding surprising results. In 2003, Bowden and hydrologist Michael Gooseff from the Colorado School of Mines in Golden began tracking the course and nutrient flow of subterranean water percolating along streambeds. They discovered that this submerged waterway and its surroundings—called the hyporheic zone—play a bigger role in stream ecology than had previously been thought. "We suspect that a major part of the nutrients to supply primary productivity actually come from [this] zone and not from the [nutrients] that run off the landscape directly," says Bowden.

Scenic science. With Alaska's Brooks Range as a backdrop, Toolik research station stands out as a center for studying global change in the Arctic.

The researchers also are finding that the hyporheic zone holds steady at 2° to 3°C below zero, and the ice above gets no colder despite air temperatures of minus 50°C. "So it only took a little bit of exposure to sunlight" to start the water flowing again, says Bowden. "The whole system is primed to go" as soon as the sun returns to the sky, he adds.

Opportunities and threats

Despite their concentrated efforts each summer, scientists are concerned that they might be missing part of the climate change story. "We are basing what we know [about the tundra] on data from June and July," says Brian Barnes, director of the University of Alaska Institute of Arctic Biology, Fairbanks, which runs the Toolik station. "We've been assuming that nothing is happening in the winter because it's too cold." But work by researchers such as Bowden is showing that underground temperatures may actually be mild enough to allow organisms to function and be on call for the spring.

Nutrients matter. A comparison of fertilized (*above*) and unfertilized (*inset*) greenhouse plots shows the importance of nutrients.

The lab is also trying to add educational components to its research agenda. Crumrine is part of an NSF-sponsored program to provide teachers with research experience in the Arctic. There's talk about starting a graduate student summer course. Also, Robin Bingham, an evolutionary biologist from Western State College in Gunnison, Colorado, is part of a group planning an Arctic biology course for undergraduates.

But such activities require additional resources, which are in short supply. The future of Toolik is closely tied to the future of NSF funding. The agency's overall budget was reduced this year and seems unlikely to do better than inflation in 2006. Forest ecologist Henry Gholz, who manages the LTER program, says both his budget and funding for ecosystems research will "likely remain static or only have a small increase." Another problem is that LTER sites are judged by how well they leverage funding from other sources. But as Hobbie notes, "in the Arctic there are no other agencies to which we can apply for funds."

Within NSF, there is increased competition for funding within programs that support work done at Toolik station. Over the past 5 years, the success rate for ecosystems proposals has fluctuated between 18% and 14%, says NSF's Michelle Kelleher. Success rates for the Office of Polar Programs have gone up and down as well: In 2000, it was 37%, but in 2005 it was 31%. The office is helping to sponsor the International Polar Year in 2007, a global effort to stimulate more research in the Arctic and in Antarctica. Without more money, Pyle says, any initiatives for the polar year will have to be paid out of the

CREDITS: E. PENNISI/*SCIENCE*

Published by AAAS

same budgets as for Arctic and Antarctic research and logistics.

There are non-fiscal threats as well. A natural gas pipeline, if it's built, would mean more people, more traffic, more way stations, and more gravel excavation. One of Shaver's sites is right next to an old gravel pit that, if reactivated, could destroy the site either directly or by increasing silt and other runoff sufficiently to invalidate longitudinal studies.

To counter these possible problems and more, Barnes and his colleagues at the University of Alaska are beginning to seek support from federal and state officials for a 44.5-hectare research park that would protect the study plots against potential intrusions. The U.S. Bureau of Land Management leases 10.8 hectares to the University of Alaska Institute of Arctic Biology as the station's grounds and has zoned the 31,000 hectares around Toolik Lake as a Research Natural Area. Expanding the size of the protected zone to include the upper Kuparak River watershed, a site of some long-term studies, would safeguard research without impeding oil and natural gas development, says Barnes.

It would also protect Toolik's future and avoid, in Gholz's words, NSF's having made "a huge investment that's thrown out." Toolik deserves special attention, Bingham and others would argue, because of its ability to monitor a key component of global climate change. "Arctic ecosystems are some of the most endangered habitats and organisms on earth," she says. **–ELIZABETH PENNISI**

Archaeology

Unraveling Khipu's Secrets

Researchers move toward understanding the communicative power of the Inca's enigmatic knotted strings, which wove an empire together

In 1956, Peruvian archaeologists uncovered a vessel hidden in the floor of a high-status home in the Inca administrative center of Puruchuco, near present-day Lima, Peru. Inside, they found a kind of treasure: a set of 21 of the knotted strings called khipu. The Inca relied on sets of khipu (or quipu in Spanish) to keep records of their far-flung realm, which extended more than 5500 kilometers, the distance from Stockholm to Cairo.

The Spanish who conquered the empire discovered that it was held together by a highly efficient bureaucracy that controlled the distribution of labor, goods, and services, using streams of khipu to issue orders and record the results. So essential were khipu to the native population, according to Galen Brokaw, an expert in Andean texts at the State University of New York at Buffalo, that the early colonial government reluctantly approved their continued use until they could be displaced by alphabetic texts the Spaniards could understand. Today, only perhaps 600 pre-Hispanic khipu survive.

For more than a century, researchers have sought to understand how these distinctive objects were used within the empire, and whether they functioned as a unique kind of three-dimensional, textile-based "writing." On page 1065 of this issue, anthropologist Gary Urton and mathematician-weaver Carrie J. Brezine, both at Harvard University in Cambridge, Massachusetts, take a step toward answering both questions. Through a computer-aided analysis of seven of the Puruchuco khipu, Urton and Brezine have identified one way that data and instructions were passed up and down the hierarchy from local villages to the powerful central government in Qosqo (modern Cusco). In the process, they also have tentatively made the first-ever identification of a khipu "word."

Almost simultaneously, archaeologist Ruth Shady Solis of the National University of San Marcos in Lima has independently unveiled what is seemingly the oldest khipu—or, perhaps, proto-khipu—ever discovered. Found in a cache buried inside a pyramid at Caral, an ancient city north of Lima that Shady's team has been excavating since 1994 (*Science*, 7 January, p. 34), the object resembles an Inca khipu, except that the pendant strings are twisted around small sticks.

First strings. This artifact from the ancient city of Caral may be a khipu as old as 4500 years.

CREDIT: PILAR OLIVARES/REUTERS

According to Shady, it is more than 3000 years older than the oldest previously known khipu, which date from the 9th century C.E. If so, then khipu, though younger than the world's first writing systems of Sumerian cuneiform and Egyptian hieroglyphics, arose in the third millennium B.C.E. and are among humankind's oldest means of communication.

The Caral artifact's apparent great age of 4000 to 4500 years "indirectly strengthens the case" that the khipu were "more than numeric," notes Daniel H. Sandweiss of the University of Maine in Orono. Ancient writing methods such as cuneiform evolved over many centuries from accounting records, as scribes invented symbols to identify what was being counted. "If what Ruth has found really is a khipu ancestor," Sandweiss says, "then khipu would be following the pattern of other writing systems."

Inca khipu consist of a main cord from which dangle as many as a thousand smaller strings, the latter of which contain clusters of knots. In the 1920s, Leland Locke, an amateur scientist, argued that khipu were simply lists of numbers, with individual knots representing digits and groups of knots on a strand representing successive powers of 10. (Blank spaces function as zeroes.) Locke's rules held true for many khipu, and his view of them as mnemonic devices largely held sway until the 1970s, when the Cornell University husband-wife team of Robert and Marcia Ascher overhauled his work, assembling a detailed khipu database (http://instruct1.cit.cornell.edu/research/quipu-ascher/). They argued that khipu were more akin to writing—and indeed that about 20% of surviving khipu do not fit Locke's rules.

If khipu were a form of writing or proto-writing, they were unlike any other. Scribes "read" the khipu by running their fingers along the strings, sometimes while manipulating small black and white stones—in striking contrast to other cultures' ways of recording symbols, which involve printing or incising

marks on flat surfaces. "The Spaniards were bewildered by them," Urton says. "Four hundred years later, we aren't much better off."

The Aschers sparked a new push to decode khipu. Supported by a National Science Foundation grant and a MacArthur Prize, Urton and Brezine in 2002 began assembling a more sophisticated khipu database that permitted complex searches (http://khipukamayuq.fas.harvard.edu/). Among the first khipu they entered was the set unearthed at Puruchuco. According to anthropologist Carol Mackey of California State University, Northridge, these khipu were found in the home of a *khipukamayuq*, an elite scribe who created and read the khipu that recorded the flow of goods, labor, and taxes within the empire. Mackey noted that two of the khipu were almost identical—an observation that tallied with the Inca writer Guaman Poma's 1609 claim that *khipukamayuq* made multiple copies of each khipu so that "no deception could be practiced by either the Indian tribute payers or the official collectors."

Brezine realized that the pattern of string colors on the two matching khipu also "was very similar—you had sequences of four strings, each with [the same] repeated pattern of four string colors." Brezine then asked the database to identify khipu with a "similar kind of arrangement of four string colors in repeating sets." By interfacing the values in the Harvard khipu database with the popular software Mathematica, Urton says, "she was able to ask, 'Is there any instance of these strings whose sum is found on another khipu?' "

The answer was yes. Brezine's data-sifting revealed a hierarchical pattern involving seven of the 21 khipus. The hierarchy consists of three levels, each with two khipu. (Urton and Brezine removed one of the level 2 khipu, which has disappeared from the Puruchuco museum, from their analysis). Khipu on each level have identical or nearly identical number values and string colors—"the checks-and-balances aspect" of khipu accounting. And the values on lower-rank khipu add up to the values on subsections of higher-rank khipu. Thus, Urton and Brezine argue, the seven khipu represent either demands, probably from the provincial governor, for labor or goods, which lower-level functionaries broke down into components, or reports of tribute from the bottom being aggregated on their way up the ladder. Either way, Urton says, "you see how information might have been funneled upward and dispersed downward"—an essential task in controlling the large, diverse, and populous empire.

Notably, some of the cords in the level 2 khipu are only approximate sums of the corresponding cords in level 1. But the top level khipu is much more precise, with only 2 inexact totals. That suggests to Urton that "some data-manipulation was going on." The *khipukamayuq* may have been matching real figures for labor taxes on the bottom to ideal requests from the top, for example.

Line by line. A set of khipu found together (one from the set, *above)* may help in understanding khipu such as this 1200-year-old one (*top*).

CREDITS (TOP TO BOTTOM): W. CONKLIN; G. URTON

The khipu on the two top levels have introductory segments of three figure-eight knots on three strings. To Urton and Brezine, the knots on these khipu, which presumably would have circulated out of their place of origin and perhaps as far as the capital, most likely served to identify their place of origin, the palace at the place now called Puruchuco. If so, then the introductory segments give its name—the first-ever precisely deciphered "word" in khipu "writing."

"The identification seems logical to me, though we are being cautious about it," Urton says. Aware that the decoding of both Mayan and Egyptian hieroglyphics began by identifying place names, he believes that "if khipu can be deciphered, this is the kind of approach that will do it."

Urton has previously argued (*Science*, 13 June 2003, p. 1650) that khipu were a kind of binary code, with the 0s and 1s being the either-or choices faced by khipu-makers (right or left direction for knots, spin, and ply, for example). With other researchers, Brokaw has criticized this binary theory, because, he says, "there is no way to reconcile it with the decimal code in which the khipu [also] clearly participate," and because he believes it is not supported by ethnographic data. But Brokaw calls the current work "fascinating," noting that it does not directly depend on the earlier binary theory.

The increased belief that the khipu were a complex means of communication is coupled with growing recognition of the extraordinary role of textiles in the precolonial Andes. "Textiles are important to every society," says William J. Conklin, an architect and archaeologist who is a research associate at the Textile Museum in Washington, D.C. "But their role in Andean societies as carriers of meaning and power is different from anything else that I know." Conklin notes, for example, that very early textiles from Huaca Prieta, a north coast site dated to about 1500 B.C.E., were apparently not used for clothing. The "incredible fact," in Conklin's view, is that "weaving was invented for what we might call 'conceptual art'—to communicate meaning—and only afterward was it used for clothing."

Khipu, Conklin says, were part of this tradition, as possibly shown by the Caral proto-khipu. Consisting of a ladderlike assemblage of 12 cotton strings, some knotted, that are wrapped around sticks, the object was found in a sealed room within one of the large pyramids at Caral earlier this year. Along with the other objects in the cache—including pristine baskets, mysterious spheres of fiber, and what looks like netting—the apparent khipu will be displayed at a Caral exhibit in Lima's Museo de la Nación until 31 August. Shady reports that her group "soon" will submit for publication the results from a "study of the context and the material within the cache." Sandweiss cautions that the huge temporal gap between the Caral object and the earliest firmly dated khipu—one carbon-dated by Conklin to between 779 and 981 C.E.—is "puzzling." Clearly, he says, "there is a great deal more to be learned here."

–CHARLES C. MANN

Random Samples

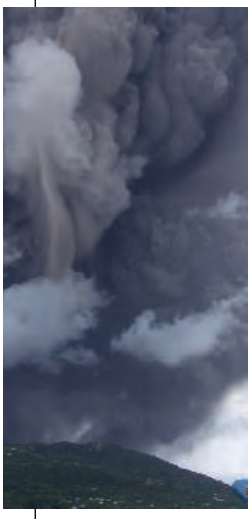
Edited by Adrian Cho

Be Fruitful and Divide

***Mirza zaza* is a new species of lemur.**

Without spotting a single new animal, researchers have identified five new species of lemurs—small nocturnal primates from the forests of Madagascar. The behavior, morphology, and genetics of the two known populations of the giant mouse lemur suggest they are in fact separate species that diverged 2 million years ago, Peter Kappeler and colleagues at the German Primate Center in Göttingen conclude in the current issue of *Primate Report*. Genetic analysis also revealed that the researchers' control group of lemurs, known simply as mouse lemurs, likewise contained a distinct new species. The two new species joined three new species of lepilemurs, also discovered by studying animals thought to be the same species and described on 10 August at the Congress of the European Federation of Primatology in Göttingen. Kappeler says the new classifications will help explain speciation among these rare animals.

Birthday Blowout



Montserrat—In an ominous birthday surprise, a slumbering volcano awakened even as researchers gathered to commemorate the 10th anniversary of its first eruption in centuries. In July 1995, the Soufrière Hills volcano on the Caribbean island of Montserrat exploded to life. Over the next few years, it rendered two-thirds of the island uninhabitable, forced thousands into exile, and killed 19 people. On 25 July, as nearly 100 researchers gathered here at a conference to mark the anniversary of the volcano's rebirth, the capricious peak suddenly belched roiling clouds of gray ash and gases into the azure sky.

The volcano had snoozed for nearly 2 years, so the renewed rumblings, which began in June, have islanders on edge. "We thought we could start moving back to some of the houses we abandoned," says John Wilson, Montserrat's minister of public works and communications. "Now I'm afraid it's going to start all over again." However, predicting what the volcano will do is difficult, says Sue Loughlin, a geologist and director of the Montserrat Volcano Observatory. And although the ash falls are annoying, she says, they are not dangerous as long as masks are worn when sweeping them up.

Ministry Scorches Italian Climate Researchers

Officials at the Italian environment ministry must like the heat. On 20 June in Rome, the ministry held a public meeting featuring climate-change skeptics who concluded that human activity plays little role in global warming. The meeting was summarized in a four-page advertisement, paid for by the ministry, in the 24 July issue of the Catholic weekly *Famiglia Cristiana*. The ad has ignited the fury of Italian climate researchers, who say they were not invited to the seminar.

According to the ad, Paolo Togni, director of communications for the environment ministry, called for an approach to environmental issues that would recognize that human activity can benefit the environment. Togni did not respond to requests for further comment. The ad also quotes Fabio Pistella, president of the Italian research council (CNR), as saying, "The phenomena of current climate changes are almost certainly natural and not due to man."

The ministry is "trying to discredit the scientific community," says Franco Miglietta, an ecologist at the Biometeorology Institute in Florence. More than 70 scientists have signed a statement calling for an open scientific debate on climate change. Meanwhile, the ministry is planning 15 more meetings on environmental issues and says it has bought advertising space for all of them.

New Age for Stonehenge Research

Despite centuries of investigation, the mystery of who built Stonehenge and why remains unsolved. Even a precise date for the prehistoric complex of stone circles, ditches, and burial mounds on England's Salisbury Plain has eluded researchers, although it's known that it was built between about 3000 and 1500 B.C.E. Now, archaeologists have proposed a new research framework to answer the many unresolved questions surrounding Stonehenge.

Much of the archaeological research carried out on Stonehenge in the 19th and early 20th centuries was of poor quality, says David Miles, chief archaeologist at English Heritage, the British government's advisory body for historical sites. This month, the organization released a report that outlines gaps in current data and proposes ways to fill them. For example, laser scanners could reveal carvings hidden by weathering and lichen, says David Batchelor, an archaeologist with English Heritage.

English Heritage has not put a time frame or price tag on the proposed research and hopes individual researchers and organizations will drum up the funding. Archaeologists may never know why Stonehenge was erected, as its builders left no written and few pictorial records of their motives, says Batchelor: "It's always going to be a bit of a mystery."

CREDITS (TOP TO BOTTOM): DAVID HARING/DUKE UNIVERSITY PRIMATE CENTER; ©GOVERNMENT OF MONTSERRAT/MONTSERRAT VOLCANO OBSERVATORY; PHOTO USED BY PERMISSION OF THE DIRECTOR, MVO; ENGLISH HERITAGE

Published by AAAS

Edited by Yudhijit Bhattacharjee

JOBS

The show goes on. Legendary computer pioneer Alan Kay is one of 14,500 employees just let go by Hewlett-Packard (HP) in a downsizing move that will terminate four research and development projects. But the pink slip isn't going to interrupt his research, Kay says.

An early architect of personal computing who created the first laptop and helped design early versions of overlapping screen windows and the Internet, Kay joined HP Labs in 2002 to lead its Advanced Software Research team in developing new software platforms based on open source code. Many of his early innovations were originally designed to reinvent computers as child-friendly, interactive educational environments, and Kay says he plans to continue that work through his California-based nonprofit organization, Viewpoints Research Institute. "We've had our agenda in place for the last 35 years," Kay says. "[Viewpoints is] a way of maintaining continuity."

HP spokesperson Dave Berman says despite the cuts, the company is still committed to innovative research in areas including printing and imaging, IT and nanotechnology, and quantum computing.

IN THE COURTS

Delayed dividend. An engineer at the University of Texas (UT), Arlington, and his former graduate student are finally getting their reward for inventing a technology that enables users of hand-held electronic devices to type words quickly. Under a $1.8 million settlement in a patent-infringement lawsuit filed by the university against the Canadian manufacturer of Blackberry devices, UT's George Kondraske (above) and Adnan Shennib, now president and CEO of the Center for Medical Device Innovations in Dublin, California, will split roughly $550,000 between them. UT Arlington will receive a $550,000 unrestricted research grant. The rest of the money from last month's settlement will go toward lawyer fees and other expenses, says UT spokesperson Anthony de Bruyn.

Kondraske and Shennib's invention allows electronic devices to deduce whole words if just a portion of the word is entered—for example, by converting "univ" to "university." This spring, 18 years after the invention, UT filed suit against more than 40 companies it charged were using the technology without paying licensing fees. Blackberry maker Research in Motion in Waterloo, Ontario, is the first to settle, although the company maintains that its own technology was developed "completely independently." In exchange for its payment, Research in Motion gets licensing rights to UT's patented technology. "The reward … comes with a sense of professional satisfaction," says Kondraske.

EXPLORERS

Then and now. Archaeology students at Michigan State University (MSU) in East Lansing didn't have to travel to exotic lands this summer for their field studies: They excavated the site where the university's first dormitory burned down in a fire 129 years ago. Besides unearthing artifacts of student life from a bygone era, including a water pitcher, a toothbrush, and clay pipe stems, the diggers confirmed archival records indicating that the accident had been caused by a fire left burning in the basement of the three-storied building by maintenance workers. "It was fascinating to discover pieces of the past lying just below our feet," says MSU anthropologist Jodie O'Gorman, one of the faculty members who led the project.

PIONEERS

A shot at the moon. Telecom inventor Sam Pitroda, who transformed India's telephone system 2 decades ago with exchanges rugged enough for rural India, now has a new mission. As chair of India's new Knowledge Commission, which the government launched last week, Pitroda aims to help India emerge as a hub for innovation and knowledge generation.

Under his leadership, the eight-member panel will chart out policy initiatives to foster excellence in India's educational institutions and public-funded research laboratories, which many analysts say are in dire need of reform. This is the second time that Pitroda, currently chair of the U.K.-based company WorldTel, has been charged with a national initiative; in the 1980s, after a successful career in American industry, he returned to India to set up the Center for Development of Telematics (C-DOT), which designed and built rural telephone exchanges. The C-DOT technology now supports more than 20 million telephone lines across India. "This new assignment is clearly a tougher challenge," says Pitroda, an electrical engineer who holds nearly 50 telecommunication patents.

The government wants the commission to come up with bold proposals to infuse creativity into Indian universities and research centers and reverse what Indian Prime Minister Manmohan Singh has described as "falling standards." Pitroda promises that the commission's first action plan, due in October, will contain some radical ideas. "In India, you have to aim at the moon to touch the roof," he says.

Got any tips for this page? E-mail people@aaas.org

CREDITS (TOP TO BOTTOM): MICHIGAN STATE UNIVERSITY; NAE; UNIVERSITY OF TEXAS AT ARLINGTON; AP

LETTERS

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 6 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

Using Ethics to Fight Bioterrorism

M. A. SOMERVILLE AND R. M. ATLAS ARE TO BE commended for drawing attention to the ethical responsibilities of life scientists whose work could impact bioterrorism ("Ethics: a weapon to counter bioterrorism," Policy Forum, 25 Mar., p. 1881). A point not given sufficient emphasis is the "professionalization of ethics," whereby a profession's commitment to a set of values and ethical standards reflects its interpretation of the world (*1*). This phenomenon tends to narrow the scope of moral evaluation to the exclusion of the experiences and needs of those who may be affected by the professional behavior (*1*). As a result, the profession's view of its role and responsibilities as embodied in its code of ethics risks becoming divorced from broader social values. To counter this tendency, it is critical that researchers engage nonscientists in the process of developing a code of ethics. A provision that acknowledges the critical role played by the public in developing a code of ethics for the life sciences (i.e., that obligates researchers to reach out to nonscientists when drafting the code) should be included together with a parallel provision that requires that the code be widely disseminated so that all those potentially affected by the research can have a basis for evaluating the conduct of scientists.

> "[I]t is critical that researchers engage nonscientists in the process of developing a code of ethics."
>
> –FRANKEL

MARK S. FRANKEL

Scientific Freedom, Responsibility and Law, American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005, USA.

Reference

1. M. Frankel, *J. Business Ethics* **8**, 109 (1989).

IN THEIR POLICY FORUM "ETHICS: A WEAPON to counter bioterrorism" (25 Mar., p. 1881), M. A. Somerville and R. M. Atlas argue convincingly that physicians and scientists in the life sciences should adopt a code of ethics against bioterrorism and bioweapon research. They propose a code of ethics that urges physicians and scientists to "[c]all to the attention of the public, or appropriate authorities, activities (including unethical research) that there are reasonable grounds to believe are likely to contribute to bioterrorism or biowarfare." This clause begs the question: What if one's own government, which presumably represents the "appropriate authorities," funds and conducts bioweapons research?

Most biodefense-related research is paid for and done at the behest of national governments, rather than by private companies or terrorist groups (*1*, *2*). The "public" and "appropriate authorities" in the code of Somerville and Atlas should be amended to "national or international public" and "appropriate national or international authorities." Because bioterrorism and bioweapon proliferation are international issues, I believe that it would be entirely appropriate for any scientist to appeal to international institutions and world opinion, if he/she has reasonable grounds to believe that his/her government is engaged in activities that are likely to contribute to bioterrorism or biowarfare. Only by cooperation and mutual supervision among nations will we have a realistic chance of limiting bioterrorism and biowarfare.

LEWYN LI

Department of Biological Sciences, Columbia University, New York, NY 10027, USA.

References

1. L. R. Ember, *Chem. Eng. News* **83**, 25 (2005).
2. S. M. Block, *Am. Sci.* **89**, 28 (2001).

M. A. SOMERVILLE AND R. M. ATLAS ("ETHICS: a weapon to counter bioterrorism," Policy Forum, 25 Mar., p. 1881) are right to call for a code of ethics to govern the conduct of research in the life sciences, but those who heed their invitation should pause before the wheel is reinvented. In listing bodies that have spoken out about a need for ethics in the conduct of life sciences research, the authors have overlooked the Council on Ethical and Judicial Affairs (CEJA) of the American Medical Association (AMA).

CEJA has addressed this need, having issued ethical guidelines to prevent malevolent use of biomedical research (*1*), which were incorporated into the AMA Code of Medical Ethics (*2*) in June 2004. These guidelines call on biomedical researchers to balance their commitment to the advancement of scientific knowledge against the same "substantive and procedural principles of ethics" articulated by Somerville and Atlas, including commitment to the betterment of public welfare and safety and the importance of maintaining public trust. Moreover, building on the scientific traditions of individual and collective responsibility, the guidelines specify that scientists should strive to assess foreseeable ramifications of their research and that of their peers in an effort to balance the promise of benefit from biomedical innovation against potential harms from corrupt or unintended application of findings.

SHANE K. GREEN*

Center for Public Health Preparedness and Disaster Response, American Medical Association, 515 North State Street, Chicago, IL 60610, USA. E-mail: shane.green@ama-assn.org

*The views and opinions expressed are those of the author and should in no way be construed as representing official policies of the American Medical Association.

References

1. E-2.078 Guidelines to Prevent Malevolent Use of Biomedical Research (available through the AMA online Policy Finder, at www.ama-assn.org/ama/noindex/category/11760.html).
2. American Medical Association Council on Ethical and Judicial Affairs, Code of Medical Ethics: Current Opinions with Annotations, 2004–2005 Edition (AMA Press, Chicago, 2004).

IN THEIR POLICY FORUM DISCUSSING THE ROLE of ethics in combating bioterrorism, M. A. Somerville and R. M. Atlas discuss the need for a code of conduct for scientists working in dual-use research areas ("Ethics: a weapon to counter bioterrorism," 25 Mar., p. 1881). For the 6th Framework Program (2002–06), the European Commission has adopted an ethics review process that considers classical ethical issues (like the use of human biological samples), human data protection, and animal testing as well as dual-use research (*1*).

Grant applicants are asked to consider dual-use aspects of their projected research (*2*). They must provide information on what dual-use implications they foresee, how to address these issues, and how relevant legal requirements will be met. The scientific reviewers of the proposals are also asked to reflect on ethical issues and to flag any sensitive areas. If the applicants and/or the scientific reviewers have indicated any ethical sensitive issues, the proposal undergoes an ethics review. An independent, multidisciplinary, multinational expert panel, including dual-use specialists, reviews the project proposal and provides recommendations to the funding institution and the applicants. These specialists are guided by EU-wide accepted international agreements, relevant EU legislation, and relevant national legislation (*3*).

Published by AAAS

The experiences gained in this established, systematic, and institutionalized review process applied by the European Commission may provide relevant information to develop a best practices model to increase awareness of the dual-use problem and to enforce ethically sound research.

JOHANNES RATH AND BERNHARD JANK
Department of Evolutionary Biology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria. E-mail: johannes.rath@univie.ac.at

References
1. See http://europa.eu.int/scadplus/leg/en/lvb/i23012.html
2. See http://europa.eu.int/comm/research/science-society/ethics/rules_en.html.
3. See http://europa.eu.int/comm/research/science-society/ethics/review_en.html.

M. A. SOMERVILLE AND R. M. ATLAS HAVE grossly oversold the use of ethics as a means of deterring the threat of bioterroism ("Ethics: a weapon to counter bioterrorism," Policy Forum, 25 Mar., p. 1881). A "Code of Ethics" similar to the one presented by the authors will not counter bioterrorism because it will be imposed on a scientific community that has no intent to conduct terror operations. The "dual-use" misnomer implies that biological scientists inadvertently develop weapons through their benign research efforts. In fact, bioterror plots are not hatched by benevolent scientists having a momentary ethical lapse; rather, they are the trade-craft of ruthless murderers who happen to have subverted the accomplishments of modern science. Bioterrorism is much more than the mere absence of ethical intentions; it is the highest crime perpetrated against humanity and is best addressed by aggressive law enforcement and intelligence efforts directed against would-be perpetrators. Imposing ethical standards on scientists as a means of curbing the use of biotechnology in terrorist plots will only demonize the scientific community.

Somerville and Atlas acknowledge that bioterrorists will not be deterred by a code of ethics. The enemy is a terrorist first and a scientist by convenience. His terroristic intentions compel him to offer his intellect to the will of his cause. The priority of cause over vocation in the many examples of scientists and doctors turned terrorist show that each was well aware of their "ethical obligations," but, undeterred by the oaths they took, chose killing over healing.

BEN PERMAN
Advanced Systems and Concepts Office, Military Operations Division, Defense Threat Reduction Agency, 8725 John J. Kingman Avenue, Fort Belvoir, VA 22060–6201, USA.

"[B]ioterror plots are ... the trade-craft of ruthless murderers who happen to have subverted the accomplishments of modern science."
–PERMAN

IN "ETHICS: A WEAPON TO COUNTER BIOTERRORISM" (Policy Forum 25 Mar., p. 1881), M. A. Somerville and R. M. Atlas cite the Thomas Butler case as a purported example of unethical behavior by a scientist. They could better have used the case to show how ethical behavior can backfire on a scientist in an age of severe political pressures on antiterrorism agencies.

Butler would not be serving a 2-year sentence had he not voluntarily reported missing vials containing plague and had he not refused on principle to plead guilty to a false accusation of lying to federal authorities as proposed in a plea-bargain offer. Somerville and Atlas state that Butler reported missing vials and then "claimed that he had inadvertently destroyed the cultures" without mentioning

Published by AAAS

that the second, contradictory statement was produced during an FBI interrogation in which Butler was in wrist irons connected to a metal waist belt, given a lie-detector test, and questioned for long hours without adequate food or sleep, and without legal counsel (waived by Butler in his wish to cooperate with authorities). Butler's self-contradiction under these circumstances led to a charge of lying (*1*) that enabled FBI officials to blame Butler for their costly actions in response to his initial report. His refusal to accept that charge led to the piling on of 54 additional charges (*2*), most derived from a dispute with his university over research contracts. The jury acquitted Butler of lying, the judge (*3*) praised his humanitarian and ethical behavior, and his former dean (*4*) has defended his use of funds.

Protests by scientists (*5*) do not imply "acceptance of [Butler's] breach of laws and regulations" as stated by Somerville and Atlas. A Code of Ethics for scientists and scientific institutions needs to reflect a better appreciation of the complex issues involved than shown by Somerville and Atlas in their discussion of the Butler case.

PETER C. AGRE,[1] SIDNEY ALTMAN,[2] ROBERT F. CURL,[3] TORSTEN N. WIESEL,[4] MARY JANE WEST-EBERHARD[5]

[1]Department of Biological Chemistry, Johns Hopkins School of Medicine, 725 North Wolfe Street, Baltimore, MD 21205–2185, USA. [2]Department of Molecular, Cellular and Developmental Biology, Yale University, 219 Prospect Street, New Haven, CT 06520–8103, USA. [3]Department of Chemistry, Rice University, MS 60, 6100 Main Street, Houston, TX 77005–1892, USA. [4]Rockefeller University, 1230 York Avenue, New York, NY 10021, USA. [5]Smithsonian Tropical Research Institute, c/o Escuela de Biología, Universidad de Costa Rica, Costa Rica.

References

1. Criminal complaint against Thomas C. Butler, U.S. District Court, Northern District of Texas, 15 January 2003 (available at http://news.findlaw.com/cnn/docs/bioterror/usbutler11503cmp.pdf).
2. Superseding Indictment, United States of America v. Thomas Campbell Butler, MD, United States District Court for the Northern District of Texas, Lubbock Division, filed 14 August 2003.
3. Transcript of sentencing hearing in the Thomas Butler Case (available at www.fas.org/butler/sentence.html).
4. B. E. Murray *et al.*, *Clin. Infect. Dis.* **40**, 1644 (2005).
5. P. Agre, S. Altman, R. Curl, T. Wiesel, "Statement regarding the case of Thomas Butler, Lubbock Texas" (Federation of American Scientists, Washington, DC, 2003) (available at www.fas.org/butler/nobellet.html).

Response

WE PROPOSED A CODE OF ETHICS TO COUNTER bioterrorism hoping it would stimulate dialog. As the Letters presented here attest, the code has caused people to think about the value of ethics and the need to act to protect science from misuse.

We welcome Frankel's points that the values and ethical standards enshrined in a code of ethics must reflect a broader spectrum that stretches well beyond those of a given profession and that the public must have a strong voice in deciding on these values and ethical standards. Ethics requires more than scientists just acting in good personal conscience, and the same is true of them acting collectively as a profession in good professional conscience. Rather, the broadest possible range of people and institutions must be involved in ethics decision-making.

The following are some of the ways in which we recognize the need to engage people outside science and, in particular, the public, in setting values and ethical standards for the life sciences: (i) In requiring ethics review of all research, we assume that, as is now the norm, any ethics committee would have a very broad-based membership. (ii) The concept we

> **"[T]he broadest possible range of people and institutions must be involved in ethics decision-making."**
>
> –SOMERVILLE AND ATLAS

Published by AAAS

articulate, that the scientific professions hold science on trust for society, establishes that society has the final say as to what will and will not be allowed in terms of ethics. (iii) The code includes an express duty "to bring to the attention of the public or appropriate authorities" activities that are unethical or could contribute to bioterrorism or biowarfare.

As we indicated, providing for "whistleblowing" is an essential element in implementing a code. Li makes the important point that one's own government could (and would, unless procedures were put in place to avoid it) have a conflict of interest in receiving information about its own wrongdoing. In such cases, it would not be an appropriate authority to which, as the code requires, to report. Who would be appropriate must be determined on a case-by-case basis and might include a role for other national authorities or international ones, as well as the public.

We are aware of the developments in medical ethics that Green references and our code is indebted to them, but experience has shown that professional specificity of ethical requirements is needed for scientists to personally identify with them and, as a result, apply them in practice.

Rath and Jank provide an important example of the practical operationalization of a code of ethics. We can learn much from the European experiential ethics knowledge that has resulted, and it is important to identify and build on all presently existing relevant resources. In some cases, ongoing monitoring of ethics as the research evolves, a step not mentioned by Rath and Jank, is also required.

> **"Ethics is integral to science, which means that unethical science is bad science, not just bad ethics."**
>
> –SOMERVILLE AND ATLAS

Like democracy, our code will not instantiate a perfect system, but the right question is whether we are better off with it than without it. We strongly disagree with Perman that "[i]mposing ethical standards… will only demonize the scientific community," that scientists might not inadvertently "develop weapons through their benign research efforts," and that "aggressive law enforcement and intelligence efforts" are the only ways to counter bioterrorism and are not complemented by implementing ethics. Ethics is integral to science, which means that unethical science is bad science, not just bad ethics. Like all elements of good science, ethics must be intentionally included and a code helps ensure that. It assists scientists in fulfilling their ethical responsibility to help protect against the misuse of science by those who would do harm.

Given the division within the scientific community caused by the Butler case, we are not surprised by Agre *et al.*'s letter, which reminds us that good facts are essential to good ethics. Judging the ethics of certain situations can be highly complex, especially when the facts are in dispute, as Agre *et al.* say they are in the Butler case. We welcome their criticism because we recognize that engaging in ethical analysis in real cases is an exercise of power, and power must be exercised ethically. In short, ethicists must also be reminded that they must practice their profession—that is, "do ethics"—ethically. But our fundamental point remains unchanged—unless to do so would be unethical, scientists and their colleagues must recognize the responsibility to comply with agreed-upon regulations and laws even if government officials and others encourage them to do otherwise and even if the intent of the research is noble.

Published by AAAS

In conclusion, the basis on which societal-level trust is established has shifted in postmodern Western societies from blind trust to earned trust (*1*). Earning trust requires openness, honesty, and integrity. It is a continuing process that requires the sharing of information and the informed consent of those who give their trust. Scientists must develop a manifest culture of responsibility to maintain the public trust upon which science depends.

MARGARET SOMERVILLE[1] AND RONALD ATLAS[2]
[1]McGill Centre for Medicine, Ethics, and Law, McGill University, Montreal, Canada H3A 1W9. [2]Center for the Deterrence of Biowarfare at the University of Louisville, Louisville, KY 40205, USA.

References
1. J. Katz, *Silent World of Doctor and Patient* (Free Press, New York, 1984)

Madrid Center Not Quite in Limbo

IN HIS ARTICLE "MADRID HEART CENTER TO BE rescued" (8 July, p. 229), X. Bosch summarized recent news about the Spanish National Center for Cardiovascular Research (Centro Nacional de Investigaciones Cardiovasculares, CNIC). He states that "CNIC slipped into limbo in May 2004…" In this context, "limbo" seems to mean "a place or condition of oblivion or neglect" (according to *Webster's New World Dictionary*). However, many things happened within CNIC during the last year. A new administrative manager was hired by the government, and financial support to finish the building project was provided by the Ministry of Health. Scientific activity kept pace and resulted in several contributions that will hopefully stand evaluation by usual scientific criteria (see www.cnic.es/research.htm). Hence, we have trouble grasping the precise connotation of the word "limbo" as used by Bosch.

LISARDO BOSCÁ,* SANTIAGO LAMAS,*
JUAN MIGUEL REDONDO*
Centro Nacional de Investigaciones Cardiovasculares, Melchor Fernandez Almagro 3, Madrid 28029, Spain.
*Group leaders at CNIC and Staff Scientists of the Consejo Superior de Investigaciones Científicas (CSIC).

CORRECTIONS AND CLARIFICATIONS

Letters: "Issue in Indian science" by S. Byravan (22 July, p. 557). Because of an editing error, the size of the Indian middle class was given as approximately 3 million people. It is approximately 300 million.

News of the Week: "Madrid heart center to be rescued" by X. Bosch (8 July, p. 229). Salvador Moncada's subject of study was incorrectly given as nitrous oxide; it is nitric oxide.

TECHNICAL COMMENT ABSTRACTS

COMMENT ON "Molybdenum Isotope Evidence for Widespread Anoxia in Mid-Proterozoic Oceans"

H.-F. Ling, J.-F. Gao, K.-D. Zhao, S.-Y. Jiang, D.-S. Ma

Molybdenum isotope data presented by Arnold *et al.* (Reports, 2 April 2004, p. 87) do not support their claim for a 10-fold change in oxic deposition area from mid-Proterozoic to present-day oceans. Our calculation using their model shows that euxinic area comprised only 3.7% of the oxic area in mid-Proterozoic oceans, which is not consistent with widespread anoxia.
Full text at
www.sciencemag.org/cg/content/full/309/5737/1017c

RESPONSE TO COMMENT ON "Molybdenum Isotope Evidence for Widespread Anoxia in Mid-Proterozoic Oceans"

A. D. Anbar, G. L. Arnold, T. W. Lyons, J. Barling

Ling *et al.* assume that molybdenum is removed only in oxic or euxinic (sulfidic) basins and that ocean bottom waters are either oxic or sulfidic. These simple assumptions ignore the importance of settings that are anoxic, or nearly so, but not sulfidic. Our conclusions are more consistent with plausible paleooceanographic interpretations.
Full text at
www.sciencemag.org/cg/content/full/309/5737/1017d

Comment on "Molybdenum Isotope Evidence for Widespread Anoxia in Mid-Proterozoic Oceans"

Arnold *et al.* (*1*) reported Mo isotope compositions of mid-Proterozoic black shale, which were different from those of euxinic sediments in the Black Sea. On the basis of a mass balance model calculation, they claimed that "the area of oxic sedimentation in the mid-Proterozoic oceans was nearly a factor of 10 smaller than the modern value, or the area of euxinic sedimentation approached a factor of 10 larger than the modern value, or, most likely, an intermediate situation occurred involving both substantial contraction of the extent of oxic deposition and expansion of the euxinic deep ocean." Thus, they suggested that a widespread anoxia occurred in mid-Proterozoic oceans. However, this conclusion seems based on misinterpretation of their model calculation results.

According to the Arnold *et al.* calculation of total Mo removal, the fraction of Mo removal to Mn-oxide (f_{ox}) was ~30% and of Mo removal to euxinic sediment (f_{eux}) was ~70% in the mid-Proterozoic. Today, f_{ox} is ~75% and f_{eux} is ~25%. Thus, the ratio (f_{ox}/f_{eux}) changed from <0.4 in the mid-Proterozoic to 3 today. On the basis of this number, Arnold *et al.* (*1*) suggested widespread anoxia in mid-Proterozoic oceans. However, it is the ratio of the Mo fraction removed to oxic sediment to that removed to euxinic sediment (f_{ox}/f_{eux}) that changed 10-fold, rather than the Mo fractions themselves or the oxic deposition area. In fact, according to Arnold *et al.* (*1*), the fraction of Mo removal to Mn-oxide changed from 30% in the mid-Proterozoic to 75% today, a change less than 3-fold. Our calculation using their model and data shows that oxic area in mid-Proterozoic oceans was predominant relative to euxinic area.

If we let F_{ox} and F_{eux} account for fluxes (g m^{-2} $year^{-1}$) of Mo to Mn-oxide and euxinic sediments, respectively, then the amount of Mo removal to Mn-oxide on an area of A_{ox} in a unit of time is $A_{ox}\bullet F_{ox}$, and the amount of Mo removal to euxinic sediment on an area of A_{eux} in a unit of time is $A_{eux}\bullet F_{eux}$. Assuming that the fractionation of Mo isotopes between seawater ($\delta^{97/95}Mo_{sw}$) and Mn-oxide is spatially and temporally constant ($\Delta_{sw\text{-}eux}$ = 2.1) (*1*–*3*) and that Mo fractionation between seawater and euxinic sediment is small ($\Delta_{sw\text{-}eux}$ = 0.3 to 0; 0.3 is used in the following calculation) as a result of quantitative Mo removal at euxinic conditions (*1*, *4*)—the input of Mo to oceans having an average $\delta^{97/95}$Mo value of zero (*1*, *2*, *4*) and a steady-state balance of Mo input and output and isotope composition (*1*, *4*)—then Eq. 1 of Arnold *et al.* (*1*) can be written as

$$A_{eux}\bullet F_{eux} \bullet (\delta^{97/95}Mo_{sw} - \Delta_{sw-eux}) + A_{ox}\bullet F_{ox} \bullet (\delta^{97/95}Mo_{sw} - \Delta_{sw-ox}) = 0 \quad (1)$$

If we let $F_{ox}/F_{eux} = K_{ox/eux}$, which is a measure of relative Mo removal efficiency of oxic to euxinic sediments, similar to k_{ox}/k_{eux} of Arnold *et al.* [note 39 in (*1*)], then Eq. 1 becomes:

$$\delta^{97/95}Mo_{sw} = (\Delta_{sw-eux} + K_{ox/eux} \bullet \Delta_{sw-ox} \bullet A_{ox}/A_{eux})/(1 + K_{ox/eux} \bullet A_{ox}/A_{eux}) \quad (2)$$

Eq. 2 relates seawater $\delta^{97/95}Mo_{sw}$ value to the ratio of oxic area to euxinic area (A_{ox}/A_{eux}). In modern oceans, $K_{ox/eux} \leq (75\%/99.5\%)/(25\%/0.5\%) \approx 0.015$, because Mo removal to the euxinic seafloor (which covers <0.5% of the total modern seafloor) accounts for ~25% of annual Mo removal in the oceans, whereas Mo removal to oxic seafloor (which covers >99% of the total modern seafloor) accounts for ~75% of annual Mo removal (*1*, *2*, *5*–*7*). The $K_{ox/eux}$ value is assumed to be constant [see note 39 in (1)]. Figure 1 shows the relation between $\delta^{97/95}Mo_{sw}$ and the A_{ox}/A_{eux} ratio expressed by Eq. 2 using these parameters. It shows that the present-day A_{ox}/A_{eux} ratio is ~200, corresponding to seawater $\delta^{97/95}Mo_{sw}$ = 1.6‰, whereas the mid-Proterozoic $\delta^{97/95}Mo_{sw}$ of 0.8‰ corresponded to A_{ox}/A_{eux} = ~27. This means that in the mid-Proterozoic, the euxinic deposition area was only about 3.7% of the oxic deposition area. This oceanic state is not consistent with the claim of widespread anoxia, although the effect of suboxic area on oceanic Mo isotopes needs more constraints. This may also imply that only a few percent euxinic area in the oceans (e.g., the existence of several semiclosed sea basins similar to the present-day Black Sea) could produce $\delta^{97/95}$Mo values similar to those of the mid-Proterozoic black shale analyzed by Arnold *et al.* (*1*).

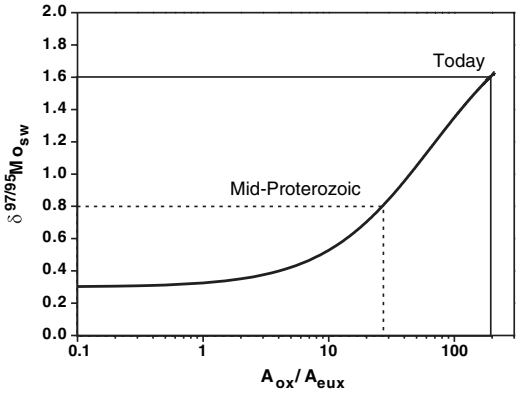


Fig. 1. Model calculation result showing that the seawater $\delta^{97/95}Mo_{sw}$ value varies as a function of A_{ox}/A_{eux} (ratio of oxic area to euxinic area). It shows that present-day A_{ox}/A_{eux} = ~200 corresponds to $\delta^{97/95}Mo_{sw}$ = 1.6‰ and that mid-Proterozoic $\delta^{97/95}Mo_{sw}$ = 0.8‰ corresponds to A_{ox}/A_{eux} = ~27.

According to this model, when $A_{ox}/A_{eux} < 10$, the variation of $\delta^{97/95}Mo_{sw}$ with A_{ox}/A_{eux} would be within analytical error (~0.24‰) (*1*), and thus A_{ox}/A_{eux} change cannot be resolved by the Mo isotopic method (Fig. 1). If at some point in geological history there were a state of widespread anoxia with euxinic area exceeding, say, 30%, then concentration and residence time of dissolved Mo in the ocean would decrease greatly. The $\delta^{97/95}$Mo value of euxinic sedimentary Mo would be close to that of input, and the oceanic Mo isotope fractionation would not be controlled by oxic sedimentation processes as today. Transition from such an anoxic condition to the oxic condition of today cannot be described using a steady-state model, as proposed by Arnold *et al.* (*1*).

H.-F. Ling
J.-F. Gao
K.-D. Zhao
S.-Y. Jiang
D.-S. Ma
State Key Laboratory for Mineral Deposits Research
Department of Earth Sciences
Nanjing University
Nanjing 210093, China
E-mail: hfling@public1.ptt.js.cn

References and Notes

1. G. L. Arnold, A. D. Anbar, J. Barling, T. W. Lyons, *Science* **304**, 87 (2004).
2. C. Siebert, T. F. Nagler, F. von Blanckenburg, J. D. Kramers, *Earth Planet. Sci. Lett.* **211**, 159 (2003).
3. J. Barling, A. D. Anbar, *Earth Planet. Sci. Lett.* **217**, 315 (2004).
4. J. Barling, G. L. Arnold, A. D. Anbar, *Earth Planet. Sci. Lett.* **193**, 447 (2001).
5. S. R. Emerson, S. S. Huested, *Mar. Chem.* **34**, 177 (1991).
6. J. L. Morford, S. Emerson, *Geochim. Cosmochim. Acta* **63**, 1735 (1999).
7. K. K. Bertine, K. K Turekian, *Geochim. Cosmochim. Acta* **37**, 1415 (1973).
8. We thank three reviewers for their comments. The authors are supported by research grants from China Ocean Mineral Resources R&D Association (DY105-01-04-6), Education Ministry of China, and Natural Science Foundation of China (40172041).

15 December 2004; accepted 5 May 2005
10.1126/science.1108737

Response to Comment on "Molybdenum Isotope Evidence for Widespread Anoxia in Mid-Proterozoic Oceans"

Relating changes in Mo isotope signatures preserved in the geologic record to changes in the extent of seafloor anoxia is not a trivial exercise. We previously presented and interpreted new Mo isotope data as indicating expanded anoxia in mid-Proterozoic oceans, potentially reaching global extent (*1*). Ling *et al*. (*2*) challenge our conclusions, arguing that the data are not consistent with widespread anoxia. We stand by our original interpretation but welcome the opportunity to clarify and expand on those conclusions, particularly in view of the potential importance of ocean redox change for biological evolution (*3*).

It is important to emphasize that Ling *et al*. accept our data and largely recapitulate our calculations to arrive at the same first-order conclusion: Within the framework of a simple two-sink model, assuming linear scaling between the fraction of Mo removed to each sink and its areal extent, the data indicate that the ratio of oxic area to euxinic area (A_{ox}/A_{eux}) was smaller by a factor of 5 to 10 during the mid-Proterozoic than today (A_{ox}/A_{eux} ~27 versus 200 today, according to Ling *et al*.). Given our present understanding of Mo isotope systematics and the available data, there is no dispute that the mid-Proterozoic Mo isotope budget was perturbed as compared with today and that this perturbation is most easily explained in terms of less oxygenated oceans at that time. The dispute is over the quantitative oceanographic implications of this shift in A_{ox}/A_{eux}.

Ling *et al*. (*2*) interpret this result to mean that because mid-Proterozoic euxinic deposition area was only about 3.7% of oxic deposition area, the claim of widespread anoxia is not valid. We reject this interpretation because the 3.7% figure describes the relative sizes of the two sinks, not the absolute size of either (although not explicitly stated by Ling *et al*., we assume that the figure was obtained as follows: $1/(A_{ox}/A_{eux}) \times 100\% = 1/27 \times 100\% = 3.7\%$). Therefore, this figure cannot uniquely constrain the areas of oxic or euxinic seafloor without an additional constraint (see below). This problem can be simply illustrated by comparison with the modern condition, $1/(A_{ox}/A_{eux}) \sim 1/200 \times 100\% = 0.5\%$ (*2*). The evolution of the oceans from a value of 3.7% in the mid-Proterozoic to 0.5% today (a roughly 7-fold change) could have occurred in one of three ways. At one extreme, A_{eux} may have contracted by a factor of ~7 since the mid-Proterozoic. Indeed, this scenario would not imply widespread anoxia in the Proterozoic; seven times the modern euxinic area would still be a trivial fraction of the total seafloor. However, at the other extreme, A_{ox} may have expanded by a factor of ~7. This scenario implies mid-Proterozoic oceans in which oxic sedimentation was vastly less common than today, fully consistent with a scenario of widespread oxygen deficiency. As stated in (*1*), it seems most likely that ocean evolution was characterized by a combination of substantial oxic contraction and euxinic expansion.

It is intuitively tempting to resolve this problem by adopting the constraint $A_{eux} + A_{ox}$ = constant. This assumption dictates that a change in the area of A_{ox} must be balanced by an equal and opposite change in the area of A_{eux}, and vice versa. Quantitatively, it demands that nearly all the change in A_{ox}/A_{eux} was the result of a 7-fold change in the size of the euxinic sink; A_{ox} changes by only 4%. We suspect that Ling *et al*. (*2*) have this constraint in mind. However, this constraint is fundamentally unrealistic because it ignores the existence of areas of seafloor other than Mn-oxide–bearing sediments and sediments underlying euxinic waters [see note 40 in (*1*)]. Such areas include suboxic settings (where bottom water $O_2 < 5$ μM but $H_2S \sim 0$ μM), settings that are anoxic but not euxinic (where $O_2 \sim H_2S \sim 0$ μM), and regions of carbonate sedimentation, detrital sediments, and the like. The existence of such areas permits A_{ox} and A_{eux} to vary independently. For example, A_{ox} could decrease independently of an increase in A_{eux} if there were an expansion of suboxic seafloor at the expense of oxic seafloor.

As discussed in (*1*), the situation deviates even further from the simple two-sink model of Ling *et al*., because evidence suggests that sediments accumulating under suboxic bottom waters are also a small but important sink for Mo; estimates suggest that this sink could account for as much as 20% of Mo cycling today (*4*). For reasons not fully understood, this sink fractionates Mo isotopes in the same direction as the oxic sink, that is, Mo isotopes in suboxic sediments are lighter than in seawater, as is the case with oxic sediments (*5*). Additionally, porewater studies in suboxic settings indicate that the net magnitude of fractionation during removal of Mo from seawater to suboxic sediments may be similar to fractionation by the oxic sink (*6*–*8*). As a result, in contrast to increases in A_{eux}, increases in the area of suboxic seafloor at the expense of A_{ox} have little or no impact on the Mo isotope system. It is difficult to imagine real ocean basins in which A_{eux} increases and A_{ox} decreases substantially without a substantial increase in A_{sub}. Therefore, we concluded that mid-Proterozoic f_{ox}/f_{eux} was less than 0.4 rather than equal to 0.4 (*1*), a finding that would translate here into $A_{ox}/A_{eux} < 27$, not $A_{ox}/A_{eux} = 27$. In fact, the Mo isotope data are consistent with mid-Proterozoic oceans in which A_{ox}/A_{eux} approached 0 and most of the seafloor was overlain by waters with $0\ \mu M < O_2 < 5\ \mu M$.

Although the efficiency with which Mo is removed to suboxic sediments is poorly known, it appears to be intermediate between the removal efficiencies in oxic and euxinic settings. Hence, it is likely that such oceans would have been typified by depressed Mo concentrations relative to today's, but not so depressed as to eliminate all authigenic enrichment of Mo. The moderate Mo enrichments in mid-Proterozoic black shales, implying ocean Mo concentrations that were depressed but within an order of magnitude of today's (*1*), fit this scenario.

Ling *et al*. conclude their comment by questioning our use of a steady-state model to describe the transition between anoxic and oxic conditions. We do not understand this critique because we apply this model to the apparently long-lived steady state between the major redox transitions of the Paleoproteozoic and Neoproterozoic. Such application of steady-state box models is not unusual.

Further study of the Mo elemental and isotope budgets will improve our ability to draw quantitative conclusions about changes in ocean anoxia from measured Mo isotope signatures. In the meantime, the possibility of widespread oxygen deficiency in mid-Proterozoic oceans, and its consequences for evolution, cannot be discounted.

A. D. Anbar
Department of Geological Sciences
and Department of Chemistry
and Biochemistry
Arizona State University
Tempe, AZ 85287, USA
E-mail: anbar@asu.edu

G. L. Arnold
Jet Propulsion Laboratory and
Division of Geological and
Planetary Sciences

California Institute of Technology
Pasadena, CA 91125, USA

T. W. Lyons
Department of Earth Sciences
University of California–Riverside
Riverside, CA 92521, USA

J. Barling
Department of Earth and Ocean Sciences
University of British Columbia
Vancouver, BC, Canada

References

1. G. L. Arnold, A. D. Anbar, J. Barling, T. W. Lyons, *Science* **304**, 87 (2004).
2. H.-F. Ling, J.-F. Gao, K.-D. Zhao, S.-Y. Jiang, D.-S. Ma, *Science* **309**, 1017 (2005); www.sciencemag.org/cgi/content/full/309/5737/1017c.
3. A. D. Anbar, A. H. Knoll, *Science* **297**, 1137 (2002).
4. J. L. Morford, S. Emerson, *Geochim. Cosmochim. Acta* **63**, 1735 (1999).
5. C. Siebert, T. F. Nagler, F. von Blanckenburg, J. D. Kramers, *Earth Planet. Sci. Lett.* **211**, 159 (2003).
6. J. McManus, T. F. Nagler, C. Siebert, C. G. Wheat, D. E. Hammond, *Geochem. Geophys. Geosyst.* **3**, 1078 (2002); 10.1029/2002GC000356.
7. A. D. Anbar, *Rev. Mineral. Geochem.* **55**, 429 (2004).
8. J. Barling, A. D. Anbar, *Earth Planet. Sci. Lett.* **217**, 315 (2004).

21 October 2004; accepted 3 August 2005
10.1126/science.1105521

NEUROSCIENCE

The Gap Between Golgi and Cajal

Edward Jones

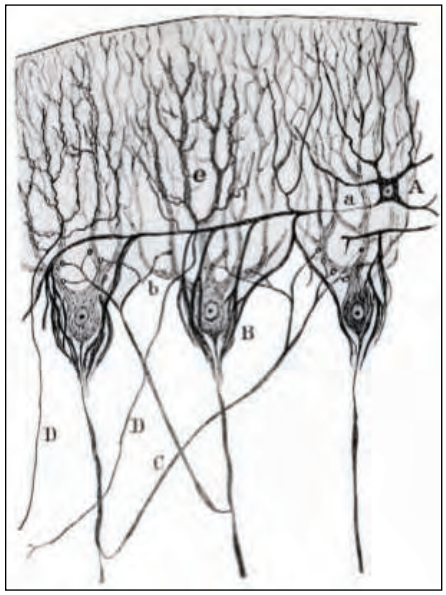
Nerve Endings
The Discovery of the Synapse
by Richard Rapport
Norton, New York, 2005. 240 pp. $23.95, C$33. ISBN 0-393-06019-5.

Few scientists are granted the privilege of controlling what posterity thinks about them. Santiago Ramón y Cajal (1852–1934) is one who was. Apart from his scientific publications, virtually all that we know or think we know about him comes from his *Recuerdos de mi vida* (*1*, *2*). Published in two volumes (1901 and 1917), it was made available to English speakers through a 1937 translation by E. Horne Craigie and Juan Cano (*3*). Their abbreviated version focuses on Cajal's description of his personal life and times but—in omitting the lengthy lists of scientific accomplishments that Cajal was at pains to document on a year-by-year basis—is of limited value to historians of neuroscience.

Cajal's most renowned scientific accomplishment, his discovery of the basis for the synapse, is the topic of Richard Rapport's *Nerve Endings*. Rapport (a neurosurgeon in Seattle, Washington) structures the story around a comparison of Cajal and Camillo Golgi. Writing for a broad, nonspecialist audience, the author interweaves the lives and research of the two scientific antagonists who shared the 1906 Nobel Prize in Physiology or Medicine.

In his *Recuerdos*, Cajal presented himself as an unruly youth who denied himself much formal education, whose medical training was crude, and whose exposure to science was nonexistent. Yet he was widely read in contemporary science and literature, and he possessed a great flair for descriptive writing that extended beyond science to the production of entertaining short stories. His knowledge of French was sufficient to permit him to prepare scientific publications and speeches in that language. Cajal was appointed a professor at age 31, by which time he was already a skilled microscopist with an enormous facility for recording in graphic form what he observed and a growing capacity to derive biological principles of fundamental importance from his observations. His patina of humility was underlain by a strong competitive streak and a desire to be recognized on the international stage.

Although Cajal worked in what he presented as an isolated environment outside the mainstream, it was not by chance that after stumbling across the Golgi technique for staining nerve cells in 1887, he was able to make the breakthrough that clinched the reality of the neuron doctrine. At the time, he was sufficiently well versed in (primarily German) scientific developments to recognize how the technique could help resolve the debate between proponents of the reticular view of the nervous system and the increasing number of researchers who saw it composed of individual cells that communicate by some means other than protoplasmic bridges. Within two years, he had shown to the world what Golgi himself (who had first described the technique in 1873) had never been able to do: When applied to the brains of small or infant animals and, in repeated impregnations, the stain consistently allowed him to show nerve cells in their entirety and to reveal their relations to one another at high resolution. The relations depended on communications between axons and somata or dendrites "by contact" and not by protoplasmic continuity. (These contacts were to become known as synapses, thanks to an unknown classical scholar who suggested the term to Charles Sherrington around 1900.) Within four years, by dint of labor remarkable for its intensity and productivity, Cajal provided all of the data that Wilhelm Waldeyer required to formally announce the neuron doctrine in 1891. He was well on the way to demonstrating the routes taken by information as it flows from cell to cell in the different regions of the central nervous system.



Nerve endings. Cajal's drawing of the terminations of stellate cell (A) axons (a), climbing fibers (D), and Purkinje axon collaterals (C) on Purkinje cells in the cerebellum from his Nobel Prize lecture, "Structure et connexions des neurones" (*5*).

CREDIT: S. RAMON Y CAJAL (5)/© THE NOBEL FOUNDATION

Cajal worked hard in promoting his science, and even at the peak of his career, he acted to ensure his priority would be recognized. While not obscuring the contributions of others, he made sure that parallel credit came his way when he and they made the same observations around the same time—as, for example, with Michael Lenhossék and the description of the growth cone, Arthur Van Gehuchten and the law of dynamic polarization, and Aldo Perroncito and the earliest outgrowth of axon sprouts in regenerating peripheral nerves. Cajal invariably gave Golgi credit for the staining technique and for the first demonstrations of the free endings of the dendrites and of the axon collaterals. Golgi, by nature far more reserved and less capable of self-promotion, grew increasingly despondent about what he viewed as his failure of recognition. The fact that he remained an adherent of the reticular theory long after its demise did not help.

Although Cajal referred to himself and Golgi as "Siamese twins joined at the shoulder," the contrast between the two scientists was strikingly revealed in Stockholm when they received their shared Nobel prize. In his Nobel lecture, Cajal in businesslike fashion presented the reality of communications by contact (he still wasn't using the word synapse) using well-chosen examples from the cerebellum, spinal cord, olfactory bulb, and retina. Then, like any modern scientist would, he quickly went on to describe the results of his latest experiments on peripheral nerve regeneration. Golgi, by contrast (and apparently to the consternation of the scientists present), devoted his lecture to an attempt to demolish the neuron doctrine. But he only succeeded in revealing himself to be out of date. In addition, his personal behavior toward Cajal may have been unpleasant—at least that is how it was reported by Cajal.

Already preoccupied with his duties as an Italian senator and with turf disputes between the universities of Pavia and Milan, after 1906 Golgi dropped further away from active research. Cajal, productive and driven as ever, continued to make new discoveries on nerve degeneration and regeneration and to undertake fresh battles over priority with the growing school of German cytoarchitectonicists, who were dividing up the cerebral cortex into functional areas. In works pub-

The reviewer is at the Center for Neuroscience, University of California, Davis, 1544 Newton Court, Davis, CA 95616, USA. E-mail: ejones@ucdavis.edu

Published by AAAS

lished in the year of his death, he was still defending the neuron doctrine against what he perceived as new threats.

Rapport retells the story of Cajal's *Recuerdos* in an engaging and entertaining way, with asides on Golgi drawn from Paolo Mazzarello's biography (*4*). Adopting the modern biographical form in which the author places himself not only in the times but also in the mind of the biographee, he writes with the faith of the true believer. *Nerve Endings* will serve to bring Cajal—who is now more often quoted than read—to the attention of a younger generation of neuroscientists. In an enjoyable fashion, it will present Cajal to them exactly as Don Santiago would have wanted.

References

1. S. Ramón y Cajal, *Recuerdos de mi vida: Mi infancia y juventud* (Moya, Madrid, 1901).
2. S. Ramón y Cajal, *Recuerdos de mi vida: Historia de mi labor científica* (Moya, Madrid, 1917).
3. S. Ramón y Cajal, *Recollections of My Life*, E. H. Craigie, J. Cano, Transls. and Eds. (American Philosophical Society, Philadelphia, 1937).
4. P. Mazzarello, *The Hidden Structure: A Scientific Biography of Camillo Golgi*, H. A. Buchtel, A. Badiani, Transls. and Eds. (Oxford Univ. Press, Oxford, 1999).
5. S. Ramón y Cajal, in *Conference Nobel, faite à Stockholm le 12 Décembre 1906* (Kungl. Boktryckeriet, Stockholm, 1907).

10.1126/science.1116404

HISTORY OF SCIENCE

A Constrained View of Evo-Devo's Roots

Manfred Laubichler

This is how one pictures the angel of history. His face is turned toward the past. Where we perceive a chain of events, he sees one single catastrophe that keeps piling ruin upon ruin and hurls it in front of his feet.…The storm irresistibly propels him into the future to which his back is turned, while the pile of debris before him grows skyward. This storm is what we call progress. —Walter Benjamin "Theses on the Philosophy of History" (*1*)

When all the hype that generally accompanies any newly proclaimed synthesis within the sciences settles, the hard work of merging diverse paradigms begins. This is the current state of evolutionary developmental biology (evo-devo). Moving beyond the catching proclamations of a "genetic toolkit" for development and the importance of "regulatory evolution" for phenotypic transformations, we are currently trying to develop a conceptual, theoretical, and empirical framework that will allow us to integrate the population-based models of microevolution with our understanding of the principles of development (genetic as well as epigenetic) and their implications for macroevolution. At such moments, the work of philosophers and historians of science can be of immense help to those scientists who are actively engaged in shaping the future of their field.

Ron Amundson, a philosopher of biology at the University of Hawaii at Hilo, has set himself a formidable task. In *The Changing Role of the Embryo in Evolutionary Thought*, he seeks to investigate the epistemological structure and ontological implications of the currently emerging synthesis of evolutionary developmental biology and to reevaluate (from a philosopher's point of view) how the history of the relation between embryology

The reviewer is at the School of Life Sciences and Center for Biology and Society, Arizona State University, Post Office Box 874501, Tempe, AZ 85287–4501, USA. E-mail: manfred.laubichler@asu.edu

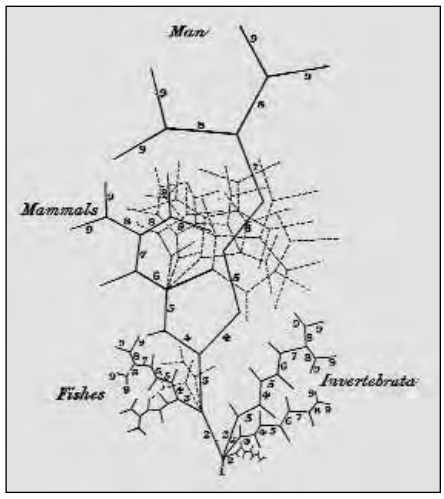
and evolutionary biology has traditionally been conceived.

The strength of Amundson's account lies in his philosophical analysis. He identifies and contextualizes four conceptual dichotomies that are deeply entrenched within the current theoretical structure of biology and that complicate any attempt to merge evolutionary and developmental biology: (i) the genotype-phenotype divide; (ii) the idea of a separation between germ line and soma; (iii) the distinction between proximate and ultimate causes, championed most prominently by the late Ernst Mayr; and (iv) the different implications of typological versus population thinking. All four dichotomies emerged during critical periods in the history of evolutionary biology, and they are now generally seen as fundamental to its conceptual structure.

As a consequence of these dichotomies, structuralist types of explanations—which involve molecular, genetic, and cellular mechanisms and for Amundson are central to developmental biology—cannot easily be fitted into the framework of population-based explanations in evolutionary biology. Amundson uses the problem of homology to illustrate the great depth of this divide. Analyzing Wagner's concept of biological homology (*2*), he concludes that the regulatory morphogenetic and morphostatic mechanisms that are at its core might offer a means of incorporating typological ideas within population thinking. But he remains skeptical, suggesting that "a true defense of evo-devo may require the refutation of the dichotomies themselves," which would be "a truly momentous achievement."

Such a defense is difficult because, according to Amundson, in the course of the second half of the 20th century the founders and advocates of the modern synthesis (foremost among them Ernst Mayr) presented a skewed perspective on the relations between developmental and evolutionary biology, one that largely created the myth of the irrelevance of development for evolution. Amundson gives an interesting account of the intellectual developments and academic politics that motivated the debates on those relations and contributed to what he sees as "synthesis historiography," a biased view of the history of evolutionary biology.

Within synthesis historiography, the history of evolutionary ideas has been reconstructed in such a way that the modern synthesis comes out "right" and that alternative accounts (such as those that emphasize architectural and developmental constraints) are discredited. With the book, Amundson wants to set the record straight, rewriting history in such a way that evo-devo comes out right and that structuralist and typological ideas are no longer dismissed. The problem is that Amundson, as he freely admits, is no historian. His typological bias, which works well for his focus on conceptual dichotomies and supports his explorations of the epistemological and methodological consequences of different explanatory frameworks, unfortunately gets in the way of writing a historical narrative that fully takes into account the complexities of historical developments. One reason for the difficulties might be that history is conceptually closer to population thinking than to typological schemas.



A developmental tree. Martin Barry's 1837 representation of von Baer's theory of the four archetypes and their developmental patterns (*6*). Darwin, in his "Notebook B" from the same year, sketched a tree of descent under which he noted "Thus genera would be formed—bearing relation to ancient types." As Robert Richards has discussed (*7*), the resemblance between the trees suggests the possibility that Darwin had learned of von Baer's embryological ideas through Barry.

The Changing Role of the Embryo in Evolutionary Thought
Roots of Evo-Devo
by Ron Amundson
Cambridge University Press, Cambridge, 2005. 294 pp. $75, £45. ISBN 0-521-80699-2. Cambridge Studies in Philosophy and Biology.

As a consequence of his approach, Amundson's version of the history of biology does contain some new and challenging perspectives. These include his discussions of the origins of essentialist ideas within biology, the concept of the fixity of species, and the latter's importance in constituting the domain of biological research in the late 18th century. But with regard to the book's principal theme—the "changing role of the embryo in evolutionary thought"—Amundson's account remains static and incomplete. Here the author has let himself be largely defined by the very same synthesis historiography that he wants to challenge. And even though he proposes to do greater justice to typological and structuralist ideas, his selection of what to discuss is constrained by the existing secondary literature in English.

Therefore he misses some of the more promising historical precedents for an evo-devo synthesis—such as Hans Przibram's program in experimental zoology (pursued in the famed Vienna Vivarium) and its emphasis on ontogenetic sequences and life history in the study of evolution (*3*) or the detailed studies in developmental physiology and physiological developmental genetics that flourished in Germany and Switzerland in the first half of the 20th century. Richard Goldschmidt is discussed briefly, but only with regard to his role as a common foe for the synthesis biologists. Amundson offers no consideration of Goldschmidt's place within a tradition that has emphasized the study of homeotic mutants and lethal factors (and thus provided the groundwork for the later discoveries of the Homeobox and the mutation screens of Christiane Nüsslein-Volhard and Eric Wieschaus). The author therefore shortchanges himself, as he does not deal with those episodes in the history of biology that would most help his case (*4*, *5*).

But does this matter? Is the history of science at all important for present scientific endeavors? Walter Benjamin's image of the angel of history is pessimistic about our abilities to consciously shape the future. Similarly, Amundson is agnostic about whether evo-devo will eventually accomplish a synthesis. Yet, as *The Changing Role of the Embryo in Evolutionary Thought* demonstrates, understanding the deep epistemological and conceptual foundations of current research practices is clearly valuable. Amundson has taken an important first step, focusing largely on conceptual and ontological incompatibilities between scientific theories, thus suggesting some order among the ruins. Because whole empires have been built on the ruins of previous ones, it pays to know more about their history; similarly in science. In the case of evo-devo, we have barely begun to sift through the debris.

References

1. W. Benjamin, *Illuminations* (Schocken, New York, 1969), chap. 10.
2. G. P. Wagner, *Annu. Rev. Ecol. Syst.* **20**, 51 (1989).
3. H. Przibram, *Experimentalzoologie* (Franz Deuticke, Vienna, 1907–1930).
4. M. D. Laubichler, H.-J. Rheinberger, *J. Exp. Zool. B.* **302**, 103 (2004).
5. G. P. Wagner, M. D. Laubichler, *J. Exp. Zool. B.* **302**, 92 (2004).
6. M. Barry, *Edinburgh New Philos. J.* **22**, 345 (1837).
7. R. J. Richards, *The Meaning of Evolution: The Morphological Construction and Ideological Reconstruction of Darwin's Theory* (Univ. Chicago Press, Chicago, 1992).

CREDIT: M. BARRY (*6*)

10.1126/science.1112990

SECURITY

Microchallenges of Chemical Weapons Proliferation

Tuan H. Nguyen

The Geneva Protocol of 1925 was established to ban the use of asphyxiating and poisonous gases in international conflict but did not restrict their possession. The stockpiling of chemical weapons continued throughout World War II and the Cold War with the development of more lethal and sophisticated agents. The continuing threat of chemical warfare and the actual use of chemical weapons in the developing world led to multilateral negotiations of the Chemical Weapons Convention (CWC), which entered force in 1997 (*1*). The CWC was drafted according to the practices of the chemical industry at the time (*2*). However, recent technological advancements within the chemical industry have the potential to threaten the effectiveness of the CWC's provisions for verification and compliance.

The CWC defines chemical weapons comprehensively and contains an unprecedented and intrusive verification mechanism, going far beyond other treaties banning weapons of mass destruction. Essential to the continued efficacy of the CWC is the provision in Article VIII for Review Conferences to "take into account any relevant scientific and technological development." The Scientific Advisory Board (SAB) for the CWC prepared a report (*3*) on relevant advancements in science and technology for the First Review Conference in 2003. Two key issues arose: newly discovered toxic agents and the development of new chemical production processes. The former has been debated extensively; the latter has not received much attention.

The International Union of Pure and Applied Chemistry (*4*) identified microprocess technology for chemical synthesis as having security implications. Research in microtechnology has grown rapidly within the last decade and has garnered much attention in both academia (*5*) and industry (*6*), with about 1000 studies detailing its possible applications (*7*). Miniaturized reaction systems for chemical synthesis and production present many advantages over traditional batch vessel methods. Microreactors have dimensions ranging from credit card size to notebook size and possess inner channels that are generally under a millimeter in diameter, thus increasing the surface area–to–volume ratio as compared with conventional chemical reactors. Industrial experts have stated (*8*), "continued development of these devices is expected to drive construction of miniature chemical plants that are inherently safe, and can operate in an explosive or hazardous regime that may be off-limits to a conventional plant and equipment."

Although the full chemical synthesis potential of microreactors is not yet clear, the syntheses of lethal chemicals (*9*) such as hydrogen cyanide, phosgene, and methyl isocyanate have been demonstrated. Nitroglycerine has recently been produced in China (*10*) by using microprocess technology at a maximum rate of 10 kg per hour.

The inherently small physical size of the equipment and small space required make it attractive for clandestine operations. The ability to produce chemicals of interest in a safer and more feasible manner, with little signature produced, could encourage their application for malicious intent. Chemical weapon precursors could be synthesized instead of purchased. "Just in time" production of chemicals could be facilitated by the use of microreactors, which would also reduce the risk of discovery and of handling and storage of dangerous and toxic chemicals in large quantities. This technology could potentially alter the expected list of signatures for chemical weapons production and thus create more challenges to nonproliferation efforts. What is the potential for widespread proliferation? How will security interests be served without hindering the technology's development and economic impact (*11*)?

Effective measures for control and verification must not curtail the development and growth of such technologies. Yet the security challenges should not be ignored. To address these issues, the Organization for the Prohibition of Chemical Weapons (OPCW) should begin by partnering not only with industry experts but also with innovators of this technology to identify and characterize immediate threats associated with these advancements.

Because the CWC reaches into a transnational, multibillion dollar chemical industry, there are limits to controls that must be defined. Transit states in Asia, where sensitive technologies exist, have only limited, unilateral export controls. The Australia Group (*12*) is an informal arrangement of 33 countries and the European Commission that aims to minimize the risk of assisting chemical and biological weapons proliferation by ensuring cooperative national export licensing. They are well positioned to begin international discussions dealing with these issues and to add microreactors specifically to their list of dual-use chemical manufacturing facilities, equipment, and related technologies.

United Nations Security Council Resolution 1540 (*13*), which was adopted in April 2004, calls on nations to adopt legislation to criminalize proliferation activities and to develop appropriate, effective export controls. It remains to be seen whether international governments will provide the necessary teeth to this resolution. However, microprocess technology is racing ahead, and dialogues must begin between the technical and diplomatic communities. The next conference to consider revisions to the CWC is scheduled for 2008. It is not too early to begin discussions to add this to the agenda.

Herbert F. York Fellow, Center for Global Security Research and Chemistry and Materials Science, Lawrence Livermore National Laboratory, University of California, Livermore, CA 94551, USA. E-mail: thnguyen@llnl.gov

References and Notes

1. Convention on the Prohibition of the Development, Production, Stockpiling, and Use of Chemical Weapons and on Their Destruction (www.opcw.org/).
2. M. Moodie, "The BWC Protocol: A critique" (Chemical and Biological Arms Control Institute, Washington, DC, June 2001), p. 17 (www.cbaci.org).
3. "Report of the Scientific Advisory Board on developments in science and technology" (RC-1/DG2 Organisation for the Prohibition of Chemical Weapons, 23 April 2003 (www.opcw.org/cwcrevcon/).
4. Lectures presented at the IUPAC Workshop: Impact of scientific developments on the Chemical Weapons Convention, Bergen, Norway, 30 June to 3 July 2002; E. D. Becker, Ed., *Pure Appl. Chem.* **74**, 2229 (2002).
5. K. F. Jensen, *Chem. Eng. Sci.* **56**, 293 (2001).
6. X. Zhang, S. Stefanick, F. Villani, *Org. Process Res. Dev.* **8**, 455 (2004).
7. K. Jähnisch, V. Hessel, H. Löwe, M. Baerns, *Angew. Chem. Int. Ed.* **43**, 406 (2004).
8. Trend reports of ACHEMA 2003, Frankfurt am Main, 19 to 24 May 2003 (www.dechema.de/8___Microreactor-lang-en.html).
9. G. Parshal, *Pure Appl. Chem.* **74**, 2259 (2002).
10. A. Thayer, *Chem. Eng. News* **83**, 43 (2005).
11. M. Hiebert, B. Dolven, *Far Eastern Econ. Rev.* **2002**, 13 (August 2002).
12. Australia group (www.australiagroup.net).
13. U.N. Security Council Resolution 1540 (www.un.org/Docs/sc/unsc_resolutions04.html).
14. Supported by the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory, under contract W-7405-Eng-48. I thank Ronald F. Lehman for contributions and comments.

10.1126/science.1112445

COSMOLOGY

Anthropic Reasoning

Mario Livio and Martin J. Rees

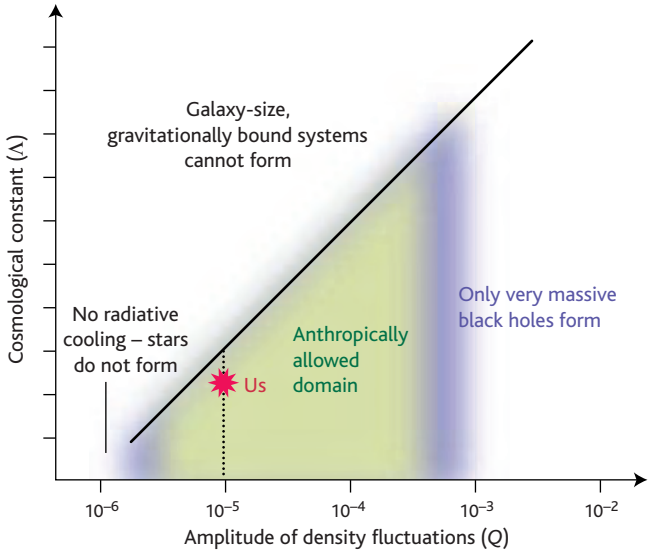
Does extraterrestrial intelligent life exist? The fact that we can even ask this question relies on an important truth: The properties of our universe have allowed complexity (of the type that characterizes humans) to emerge. Obviously, the biological details of humans and their emergence depend on contingent features of Earth and its history. However, some requirements would seem generic for any form of life: galaxies, stars, and (probably) planets had to form; nucleosynthesis in stars had to give rise to atoms such as carbon, oxygen, and iron; and these atoms had to be in a stable environment where they could combine to form the molecules of life.

We can imagine universes where the constants of physics and cosmology have different values. Many such "counterfactual" universes would not have allowed the chain of processes that could have led to any kind of advanced life. For instance, even a universe with the same physical laws and the same values of all physical constants but one—a cosmological constant Λ (the "pressure" of the physical vacuum) higher by more than an order of magnitude—would have expanded so fast that no galaxies could have formed. Other properties that appear to have been crucial for the emergence of complexity are (i) the presence of baryons (particles such as protons and neutrons); (ii) the fact that the universe is not infinitely smooth, allowing for the formation of structure (quantified as the amplitude of the fluctuations in the cosmic microwave background, Q); and (iii) a gravitational force that is weaker by a factor of nearly 10^{40} than the microphysical forces that act within atoms and molecules—were gravity not so weak, there would not be such a large difference between the atomic and the cosmic scales of mass, length, and time.

A key challenge confronting 21st-century physics is to decide which of these dimensionless parameters such as Q and Λ are truly fundamental—in the sense of being explicable within the framework of an ultimate, unified theory—and which are merely accidental. The possibility that some are accidental has certainly become viable in the context of the "eternal inflation" scenario (*1*–*3*), where there are an infinity of separate "big bangs" within an exponentially expanding substratum. Some versions of string theory allow a huge variety of vacua, each characterized by different values of Λ (or even different dimensionality) (*4*). Both these concepts entail the existence of a vast ensemble of pocket universes—a "multiverse." If some physical constants are not fundamental, then they may take different values in different members of the ensemble. Consequently, some pocket universes may not allow complexity or intelligent life to evolve within them. Humans would clearly have to find themselves in a pocket universe that is "biophilic." Some otherwise puzzling features of our universe may then simply be the result of the epoch in which we exist and can observe. In other words, the values of the accidental constants would have to be within the ranges that would have allowed intelligent life to develop. The process of delineating and investigating the consequences of these biophilic domains is what has become known as anthropic reasoning.

Anthropic considerations are beginning to be seriously discussed, especially in relation to dark energy. During the past 7 years, it has become clear that the expansion of the universe is accelerating (*5*, *6*), with dark energy contributing about 70% of the critical density required to sustain a geometrically flat universe (*7*). The question that arises is why we happen to live in the first and probably only time in the history of the universe in which the matter density and dark energy density are roughly equal.

The questions used to be: Why should empty space exert a force? Why should there be a cosmological constant Λ? Now we ask: Why is the force so small? If there was an inflationary era with a large cosmic repulsion, how could that force have switched off (or somehow have been neutralized) with such amazing precision? In our present universe, Λ is lower by a factor of about 10^{120} than the value that seems natural to theorists.



You are here. Plot of cosmological constant Λ versus amplitude of fluctuations in cosmic microwave background Q. Shaded region shows conditions that allow existence of complexity.

Following the original suggestion by Weinberg, some models have been constructed in which the cosmological constant is a random variable, whose a priori probability distribution is determined by the laws of physics (in the framework of inflationary cosmology). If we assume that we are typical (mediocre) observers, the approximate coincidence between the time when the dark energy starts to dominate the cosmic energy density and the present age of the universe does seem to find a natural explanation: If Λ were larger, then the acceleration would have overwhelmed gravity before galaxies had a chance to form (*8*, *9*).

The situation becomes more complex when more than one physical parameter is postulated to be a random variable. For instance, Q, the amplitude of the fluctuations generated in the aftermath of the big bang, could take different values in other universes (*10*). However, in the anthropically allowed domain, Λ and Q could be correlated, in the sense that in a universe that has higher-amplitude density fluctuations, structure could still form even in the presence of considerable acceleration. For Q values smaller than about 10^{-6}, only small structures of dark matter would form, and even those would do so rather late in the

CREDIT: PRESTON HUEY/*SCIENCE*

M. Livio is at the Space Telescope Science Institute, Baltimore, MD 21218, USA. E-mail: mlivio@stsci.edu
M. J. Rees is at the Institute of Astronomy, University of Cambridge, Cambridge CB3 0HA, UK. E-mail: mjr36@cam.ac.uk

Published by AAAS

universe's life (see the figure). Gas within these structures would be so dilute that it would not cool radiatively, thereby precluding star formation. For values of Q larger than about 10^{-3}, large structures would collapse gravitationally to form mostly monstrous black holes. For any given value of Q, above a certain value of Λ no galactic-mass bound systems would form before accelerating expansion commences. Clearly, our own pocket universe lies within the anthropically allowed domain. However, a more definitive answer to the question of how "typical" our location is will have to await a better understanding of the probability distributions of Λ and Q over the ensemble of universes.

We would argue that both the multiverse and anthropic reasoning are bona fide topics for (albeit speculative) scientific discourse. Improved understanding of string theory or inflation may give us firmer grounds for assessing whether our big bang was unique, or merely one of many. In the latter case, some of the physical constants and cosmic numbers that we have hitherto regarded as fundamental are in fact accidental. Yet the mere mention of "anthropic reasoning" and the "multiverse" tends to raise the blood pressure of some physicists. There are two reasons for this.

First, the potential existence of an ensemble of unobservable universes appears to be in conflict with the "scientific method" (which requires theories to be falsifiable by observations or experiments) and therefore in the realm of metaphysics. But there is actually a "fuzzy" boundary between what we define to be observable and what is not. The capabilities of present-day telescopes obviously imply a "horizon" beyond which nothing can be observed; a more fundamental limit to observations lies at the particle horizon, that is, the spherical surface around us (effectively at infinite redshift) from which photons emitted at the big bang are just reaching us.

In the simplest universe described in the textbooks, the Einstein–de Sitter model, the expansion decelerates, the redshift of any galaxy decreases, and all galaxies currently lying beyond the horizon will eventually come within it; they will thus become observable in the far future. But we are not in an Einstein–de Sitter universe: We are in a flat but accelerating one. In this universe, any galaxies now beyond the horizon will stay that way. A generic accelerating universe contains galaxies that we can never, even in principle, observe. If there are galaxies in "our" pocket universe—the aftermath of our big bang—that will never be observable, is it then much of a leap to envisage unobservable galaxies that came into being from other big bangs—part of a multiverse?

As mentioned above, there are theories that predict many big bangs: One possibility is the existence of brane worlds, that is, many universes embedded in a higher dimensional space. Another is "eternal inflation." What we have traditionally called "the universe" could be just one patch of space-time in a vast cosmic archipelago—a "landscape," some call it.

We do not know if these theories are correct. But they are speculative science, not metaphysics. Could there be many big bangs? If so, are they characterized by a range of values of the cosmic numbers—Q, Λ, and so forth? Would the same physics apply throughout the multiverse? What could give us confidence in unobservable universes?

The answer is clear—we will believe in them if they are predicted by a theory that gains credibility because it accounts for things that we can observe. We believe in quarks, and in what general relativity says about the inside of black holes, because our inferences are based on theories corroborated in other ways. Specifically, if a theory has testable and falsifiable predictions in the observable parts of the universe, we should seriously consider and be prepared to accept its predictions in parts of the universe (or multiverse) that are not accessible to direct observations.

A second reason why some physicists are hostile to the multiverse concept is that anthropic reasoning seems to point to a fundamental limitation of physics—even the "end of physics." But this objection is, in our opinion, a purely psychological one. Physicists would like, above all else, to discover a uniquely self-consistent set of equations that determines all microphysical constants, and the recipe for the big bang. They therefore hope that future theories will reveal that all physical parameters are uniquely determined. But there is no reason why physical reality should be structured according to their preferences. It is good that many physicists are motivated to seek a theory that uniquely derives all fundamental numbers and constants, but they may be doomed to failure.

The quest for first-principles explanations may prove as vain as Kepler's quest for a beautiful mathematical formula that described the solar system. If future developments bear out the possibility of a multiverse, then anthropic arguments will offer the only "explanation" that we will ever have for some features of our universe. At the moment, we have no firm reason to close off any of the options. In view of our current ignorance as to what is truly fundamental and what is not, we should keep an open mind about all the options. What we have traditionally called fundamental constants and laws could be mere parochial bylaws in our cosmic patch. They might derive from some overarching theory governing the ensemble, but might not be uniquely fixed by that theory.

Finally, one may wonder whether anthropic reasoning has any predictive power. In principle it has. For instance, imagine that the cosmological constant is the only random variable. If, as some arguments suggest, it is drawn from a flat probability distribution, then in a random member of the multiverse one would expect it to take a high value. In the subset of anthropically allowed universes, however, clearly there is some upper limit above which structure and complexity would not emerge. If our universe were a mediocre member of the ensemble (as Copernican humility would require), the expectation is that the value of the cosmological constant in our universe would not be much smaller than this upper limit. Put differently, if observations showed that the cosmological constant is smaller than the anthropic threshold by a factor of 10^5, this would make any anthropic arguments seriously questionable. As it turns out, however, the value determined from observations of high-redshift supernovae and from the spectrum of the fluctuations in the cosmic microwave background is smaller than the threshold by only a factor of 5 to 10, not inconsistent with anthropic expectations.

The next few decades are expected to witness better constraints on the nature of dark energy, experimental tests of supersymmetry and symmetry breaking, and perhaps the detection of the gravity waves that originated from inflation (*11*). These and other unexpected discoveries will undoubtedly shed some light on the reality of the multiverse and on the uniqueness (or not) of the laws of physics. Our universe isn't the neatest and simplest. It has the rather arbitrary-seeming mix of ingredients in the parameter range that allows us to exist. Until we know for sure which type of universe or multiverse we live in, anthropic reasoning is certainly one option in the physicists' arsenal.

References

1. P. J. Steinhardt, in *The Very Early Universe*, G. W. Gibbons, S. Hawking, S. T. C. Siklos, Eds. (Cambridge Univ. Press, Cambridge, 1983), p. 251.
2. A. Vilenkin, *Phys. Rev. D* **27**, 2848 (1983).
3. A. D. Linde, *Mod. Phys. Lett. A* **1**, 81 (1986).
4. S. Kachru, R. Kallosh, A. Linde, S. P. Trivedi, *Phys. Rev. D* **68**, 046005 (2003).
5. A. G. Riess *et al.*, *Astron. J.* **116**, 1009 (1998).
6. S. Perlmutter *et al.*, *Astrophys. J.* **517**, 565 (1999).
7. D. N. Spergel, *Astrophys. J. Suppl.* **148**, 175 (2003).
8. S. Weinberg, *Phys. Rev. Lett.* **59**, 2607 (1987).
9. J. Garriga, M. Livio, A. Vilenkin, *Phys. Rev. D* **61**, 023503 (2000).
10. M. Tegmark, M. J. Rees, *Astrophys. J.* **499**, 526 (1998).
11. M. Tegmark, *J. Cosmol. Astropart. Phys.* 10.1088/1475-7516/2005/04/001 (4 April 2005).

10.1126/science.1111446

PLANT SCIENCE

The Right Time and Place for Making Flowers

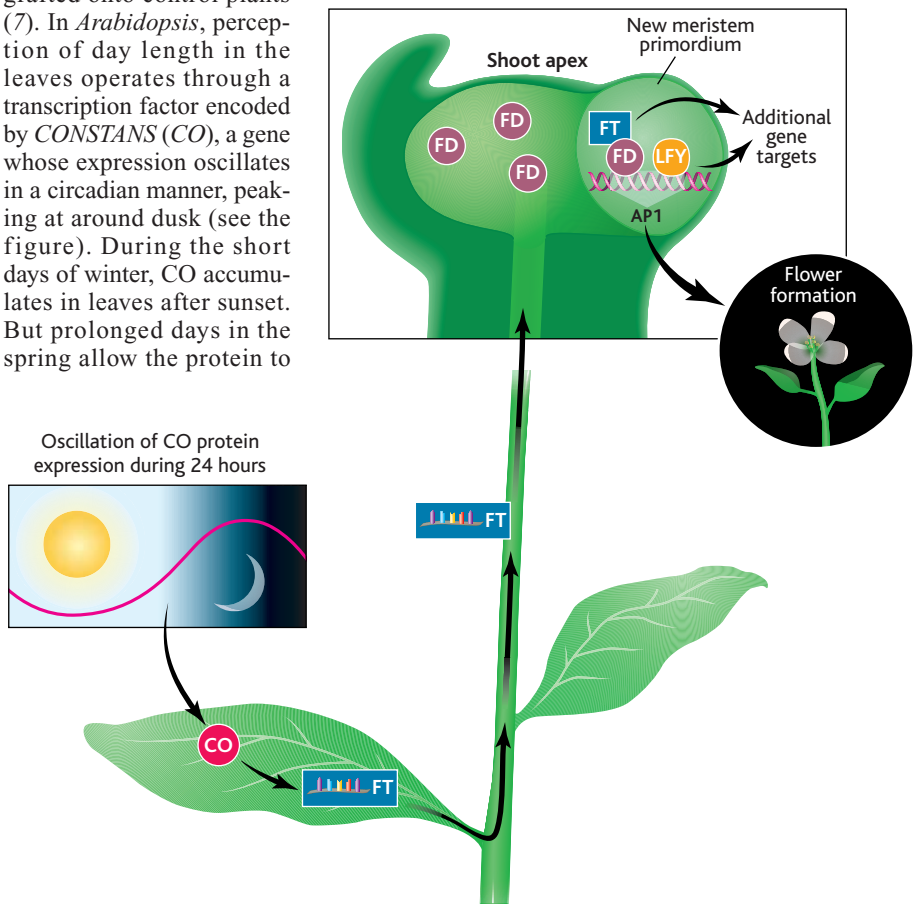
Miguel A. Blázquez

Reproductive success in plants depends on the synchronization of flowering within a given species. Many plants have developed a highly complex signaling network that monitors environmental conditions, such as day length, temperature, or nutrient availability, and determines the appropriate timing for flowering (*1*, *2*). This is the case for the model plant *Arabidopsis thaliana* and the pea that both flower in spring when day length and ambient temperature increase, or certain rice varieties and soybean that flower early in the fall when days get shorter. The initiation of flowering requires an additional developmental program to specify the floral identity of the new structures that continuously arise at the shoot apex (*3*). For instance, during the long vegetative phase in *Arabidopsis*, every primordium, the groups of cells poised to differentiate, forms a leaf. However, once the decision to flower has been made, all newly emerging primordia follow a developmental program that culminates in the formation of flowers rather than leaves. Thus, constructing a flower requires both temporal and spatial information that restricts the initiation of flowering to specific locations. But how this information is integrated has not been clear. Three studies now reveal the molecular mechanism by which this integration is achieved. In this issue, Abe *et al*. on page 1052 (*4*) and Wigge *et al*. on page 1056 (*5*) report that interaction between Flowering Locus T (FT), a protein encoded by a gene that is expressed in leaves, and FD, a bZIP transcription factor that is present only in the shoot apex, triggers the expression of floral identity genes in the new primordia. The third paper by Huang *et al*. in this week's *Science* Express (*6*) reports how the two factors meet—FT transcript travels from leaf to shoot via the plant vascular tissue.

Enhanced online at www.sciencemag.org/cgi/content/full/309/5737/1024

The author is at the Instituto de Biología Molecular y Celular de Plantas (UPV-CSIC), Universidad Politécnica de Valencia, 46022 Valencia, Spain. E-mail: mblazquez@ibmcp.upv.es

It has been known for at least 50 years that flowering is triggered at the shoot apex through a mobile signal, or "florigen," that is generated in leaves in response to conditions that promote the production of flowers. In a classic experiment, the leaves of florally induced *Perilla crispa* plants promoted flowering when grafted onto control plants (*7*). In *Arabidopsis*, perception of day length in the leaves operates through a transcription factor encoded by *CONSTANS* (*CO*), a gene whose expression oscillates in a circadian manner, peaking at around dusk (see the figure). During the short days of winter, CO accumulates in leaves after sunset. But prolonged days in the spring allow the protein to accumulate in the presence of light, the stimulus that activates CO (*8*). Overexpression of CO causes early flowering, and among the target genes directly activated by CO, two seem to be most relevant for floral induction: *FT* and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1*. Interestingly, CO acts in the phloem, the vascular tissue of plants, to activate *FT* expression in leaves in a cell-autonomous manner. This is based on the observation that CO activates *FT* expression and promotes flowering only when expressed under the control of phloem-specific promoters in the leaf, but not apex-specific promoters in the shoot (*9, 10*). These results suggest that the activity of CO is central for the generation of the mobile signal that originates in the leaf but has to be perceived in the apex to establish flowering. The up-regulation of *FT* expression by CO is required because loss of FT function prevents early flowering caused by overexpression of *CO*, whereas increasing *FT* expression causes premature flowering (*11, 12*). Thus, an important question has been how FT produced in the leaves would activate the transcription of floral identity genes, such as *APETALA1* (*AP1*), at the shoot apex.



Integration of signals to generate a flower. Appropriate day length allows the accumulation of the transcription factor CO that controls expression of *FT* in the leaf. (**Inset**) FT transcript moves through the phloem to the shoot apex where the FT protein is produced and interacts with the transcription factor FD. The complex then activates key genes such as *AP1* to start flower development. LEAFY (LFY) is a transcription factor required for *AP1* expression in wild-type plants. *LFY* expression is up-regulated by FT in the shoot apex.

Published by AAAS

To solve the spatial paradox of FT action, Abe *et al*. and Wigge *et al*. analyzed a gene encoding a new bZIP transcription factor, FD, that is expressed preferentially at the shoot apex in the region where new primordia are being generated (*4*, *5*). Multiple lines of evidence in these studies suggest a model by which FD provides the spatial framework for timely activation of flowering by FT. First, FD is required by FT to promote flowering because mutations in the *FD* gene delayed both up-regulation of *AP1* expression and the early flowering phenotype caused by *FT* overexpression. Second, although *FD* is not as efficient as *FT* in promoting early flowering when either one is overexpressed, there was synergistic interaction between them in plants that overexpress both factors. And third, FT and FD proteins interact physically, as shown in yeast by two-hybrid assays and as seen in plants by fluorescence microscopy.

How relevant is the interaction between FT and FD for the regulation of flowering? FT has no known DNA binding domain. However, constitutive expression of a fusion protein containing FT and the glucocorticoid receptor accelerated flowering in the presence of dexamethasone, a synthetic steroid that activates the glucocorticoid receptor and allows translocation of the fusion protein into the nucleus (*4*). Furthermore, a key experiment strongly suggests that FD and FT act together to activate downstream targets: Ectopic expression of FD caused up-regulation of *AP1* expression in leaves only when they were subjected to treatments that increase *FT* expression, such as transfer of plants from short- to long-day conditions (*5*).

The finding that FT and FD act together to activate reproductive development in plants fills a gap in our understanding of how temporal information and spatial constraints are integrated, but several questions remain. For instance, it is intriguing how *AP1* expression is established precisely in floral primordia, given that *FD* is more widely expressed in the shoot apex. As proposed by Abe *et al*., other proteins must restrict *AP1* expression to the correct location, and in this context, it is worth mentioning that TERMINAL FLOWER 1, a protein with strong sequence similarity to FT, is a well-known regulator of *AP1* expression that prevents AP1 from invading the central part of the shoot apex (*13*).

The model presented by Abe *et al*. and Wigge *et al*. implies that FT itself might be an important component of the elusive mobile signal that induces flowering, because *FT* is expressed in a plant tissue different from the cells in which its direct interaction with FD is needed. The study by Huang *et al*. (*6*) answers this question, showing that the transcript of FT moves from the leaf to the shoot apex. By locally inducing *FT* expression in a single *Arabidposis* leaf, the authors demonstrate that a pulse of *FT* expression in the leaf results in transport of the FT transcript to the shoot apex, and is sufficient to trigger flowering. Indeed, long-distance movement of RNAs through the phloem has been well documented in plants (*14*), but it remains to be determined if specific proteins are invloved in the transport of FT transcripts through the phloem. In a more complicated scenario, FT presence in the apex might also be the result of the activity of a different FT-induced signal moving through the phloem or from cell to cell. Movement of transcription factors through plasmodesmata, junctions that allow direct communication between the cytoplasm of adjacent plant cells, has also been described (*15*). It remains to be determined if specific proteins are involved in the transport of FT transcripts through the phloem. Although the composition of the florigenic signal is very likely complex (*16*), it seems that our understanding of this phenomenon is coming full circle.

References

1. M. J. Yanovsky, S. A. Kay, *Nat. Rev. Mol. Cell Biol.* **4**, 265 (2003).
2. G. G. Simpson, C. Dean, *Science* **296**, 285 (2002).
3. Y. Komeda, *Annu. Rev. Plant Biol.* **55**, 521 (2004).
4. M. Abe *et al*., *Science* **309**, 1052 (2005).
5. P. A. Wigge *et al*., *Science* **309**, 1056 (2005).
6. T. Huang *et al*., *Science* 11 August 2005 (10.1126/science.1117768).
7. J. A. D. Zeevaart. *Annu. Rev. Plant Physiol.* **27**, 321 (1976).
8. F. Valverde *et al*., *Science* **303**, 1003 (2004).
9. H. An *et al*., *Development* **131**, 3615 (2004).
10. B. G. Ayre, R. Turgeon, *Plant Physiol.* **135**, 2271 (2004).
11. Y. Kobayashi *et al*., *Science* **286**, 1960 (1999).
12. I. Kardailsky *et al*., *Science* **286**, 1962 (1999).
13. O. J. Ratcliffe *et al*., *Development* **126**, 1109 (1999).
14. S. Ueki, V. Citovsky, *BioEssays* **23**, 1087 (2001).
15. A. Sessions, M. F. Yanofsky, D. Weigel, *Science* **289**, 779 (2000).
16. G. Bernier *et al*., *Plant Cell* **5**, 1147 (1993).

10.1126/science.1117203

GEOCHEMISTRY

Biogeochemical Cycling of Iron Isotopes

Clark M. Johnson and Brian L. Beard

Iron is the most abundant element that engages in reduction-oxidation (redox) chemistry. The ferrous form (Fe^{2+}) is dominant in the core, mantle, and deep crust, but ferric iron (Fe^{3+}) is stable under current atmospheric conditions and hence is the stable oxidation state in most surface environments. At the same time, some of the largest fractionations in the isotopic composition of iron {commonly expressed as $\delta^{56}Fe = [(^{56}Fe/^{54}Fe_{Sample})/(^{56}Fe/^{54}Fe_{Standard}) - 1] \times 10^3$} occur between oxidized and reduced forms. Because biochemistry involves changes in redox state, this fractionation process has been a major motivation for developing this isotopic system as a means for tracing biogeochemical phenomena. In environments that contain iron in both oxidation states, the oxidized form is generally enriched in the heavy isotopes on the order of several per mil (parts per thousand, or ‰) at room temperature. This behavior is seen across all of the transition elements that have multiple oxidation states (*1*). In terms of isotopic studies of the transition elements, iron has received the most attention because of its high abundance on Earth and its prominent role in biogeochemical processes.

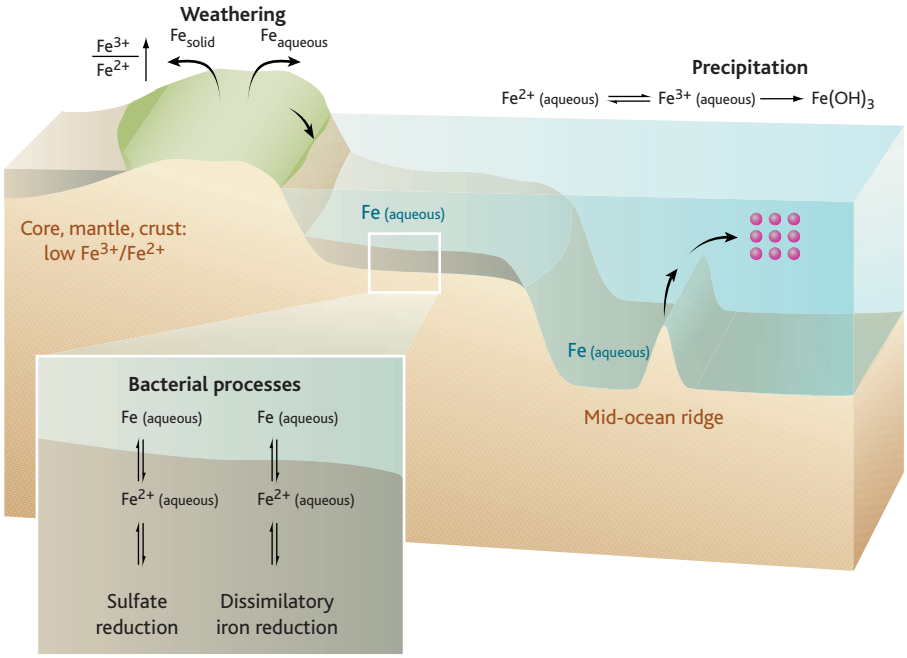
More than 60 papers have been published on iron isotope geochemistry since the field initially gained visibility in 1999, and these works have addressed issues ranging from biological processing of iron (*2*) to the rise of oxygen in the atmosphere (*3*). Collectively, studies of natural samples, as well as the critical laboratory-determined equilibrium and kinetic isotope fractionation factors in abiologic and biologic systems, have provided an initial picture of isotopic variations that are produced by global biogeochemical cycling of iron (see the figure). A remarkably large portion of the iron inventory on Earth is isotopically homogeneous ($\delta^{56}Fe = 0$‰ relative to an igneous rock standard), including igneous rocks and sedimentary rocks that have undergone minimal chemical change after deposition (*4*). Although iron-isotope variations within the mantle could arise as a result of high-pressure mineral fractionation and/or chemical changes, these variations are apparently homogenized during magma generation,

The authors are in the Department of Geology and Geophysics, University of Wisconsin–Madison, Madison, WI 53706, USA. E-mail: clarkj@geology.wisc.edu

producing oceanic and continental crust with $^{56}Fe/^{54}Fe$ ratios that are indistinguishable from those of bulk Earth (5). Terrestrial weathering, transport, and deposition produce negligible isotopic change in bulk detritus under modern atmospheric conditions, despite a large increase in the Fe^{3+}/Fe^{2+} ratio upon weathering (see the figure). Despite the large intrinsic isotope fractionation that exists between ferric and ferrous species (*6*), the low solubility of ferric iron prevents loss of a sizable proportion of iron in aqueous systems. The small proportion of soluble iron that is lost during modern weathering has relatively low $\delta^{56}Fe$ values (see the figure) (*7*, *8*). Although this mobile iron component is too small to affect the isotope composition of bulk sedimentary material, it may provide an important control on the iron isotope compositions of the modern oceans because of the very low iron content of seawater.

Weathering under a reduced (low oxygen) atmosphere, such as might have existed in the Archean, would be expected to produce much larger quantities of soluble iron, and the $\delta^{56}Fe$ values of this component should lie close to zero because no redox change would occur during weathering. Moreover, the expected greater extent of dissolution during weathering under an anoxic atmosphere would produce $\delta^{56}Fe$ values for aqueous iron that are close to the initial igneous values, as required by simple isotopic mass balance relations (*9*). An ancient, iron-rich ocean might be expected to have $\delta^{56}Fe$ values between –0.5 and 0‰, in contrast to the more negative $\delta^{56}Fe$ values that appear to be characteristic of the modern iron-poor oceans (*10*), although the relative balance of sources and sinks was likely to have been important (*11*).

Much attention has been focused on iron isotope variations in Precambrian sedimentary rocks because they may record major changes in the iron redox cycle due to changes in the contents of atmospheric oxygen, as well as development of metabolic processes that oxidize or reduce iron (*3*, *9*, *12*). The relatively large range in iron isotope compositions measured for Precambrian sedimentary rocks and minerals spans the entire range yet observed in nature ($\delta^{56}Fe$ = –4 to +1‰), and if such compositions were produced when the oceans were iron rich, extremely large quantities of iron must have been cycled. It seems unlikely that large ranges in isotopic composition over short time intervals (10^3 to 10^5 years) could directly reflect that of an anoxic, iron-rich ocean, given the great resistance such an ocean would have to changes in the isotopic compositions of sources and sinks if the residence time of iron was long (*12*).



Iron flow in the environment. Despite extensive oxidation of deep-Earth iron during weathering, no appreciable iron isotope fractionation occurs for the bulk detritus (upper left). Under oxic conditions, however, small amounts of dissolved iron may have low $\delta^{56}Fe$ values. Isotopic fractionation is important during oxidation and precipitation of hydrothermally-sourced aqueous Fe^{2+} ($\delta^{56}Fe$ ~ –0.5 ‰) through a variety of mechanisms at mid ocean ridges (at right), all of which will produce positive $\delta^{56}Fe$ values for ferric hydroxide precipitates (red spheres) through a two-step oxidation and precipitation process. The largest range in $\delta^{56}Fe$ values in the sedimentary record lies in marine sediments that have undergone anoxic diagenesis (expanded box), either in the presence of sulfide or during dissimilatory iron reduction, and these processes appear to produce distinct $\delta^{56}Fe$ values for porewater Fe^{2+}.

Oxidation of aqueous Fe^{2+} at near-neutral pH, either by oxygen or anaerobic photosynthetic iron oxidation, will produce ferric hydroxide precipitates that have positive $\delta^{56}Fe$ values (see the figure) (*13*, *14*). The very rapid isotopic exchange that occurs between aqueous Fe^{2+} and Fe^{3+} ensures that for all but essentially instantaneous oxidation rates, isotopic equilibrium will be maintained between reduced and oxidized aqueous species (*4*), suggesting that isotopic fractionation during the oxidative step will be independent of the pathway in nature, be it oxidation by atmospheric O_2, anaerobic photosynthetic Fe oxidation, or ultraviolet photo-oxidation. The $\delta^{56}Fe$ value of the final ferric hydroxide precipitates that are produced during oxidation under near-neutral pH will therefore depend on the magnitude of the Fe^{3+}_{aq}-$Fe(OH)_3$ fractionation (see the figure), which is likely to be near zero under equilibrium conditions but may be quite variable during rapid precipitation because of the unidirectional nature of the process (*15*). The overall fractionation between ferric hydroxide product and aqueous Fe^{2+} therefore lies generally between ~ +1.4 and +2.9‰ at room temperature (see the figure). Oxidation of aqueous Fe^{2+} from mid-ocean ridge hydrothermal fluids by O_2-bearing seawater, or, in the Precambrian, oxidation of Fe^{2+}-rich anoxic bottom waters that mixed with relatively oxygenated surface waters, are processes that would produce Fe^{3+}-bearing minerals with $\delta^{56}Fe > +1‰$, and yet such materials are relatively rare in the rock record (*3*, *9*).

Marine sedimentary rocks that have undergone anoxic diagenesis, from Archean age to modern environments, generally have $\delta^{56}Fe$ values less than –1‰ (*3*, *9*). Reductive dissolution of ferric hydroxide/oxide minerals during anoxic reactions may produce aqueous Fe^{2+} and Fe^{2+}-bearing minerals such as siderite and magnetite through microbial dissimilatory iron reduction (DIR), where bacteria pump electrons to the Fe^{3+} phase to create reduced iron externally. Aqueous Fe^{2+} and iron sulfide minerals also may be produced through interaction with dissolved sulfide, through abiologic sulfide pathways or by bacterial sulfate reduction. The evidence at

CREDIT: PRESTON HUEY/*SCIENCE*

hand suggests that the iron isotope compositions of aqueous Fe^{2+} that is produced by DIR ($\delta^{56}Fe$= –1.0 to –2.5‰) or sulfide interactions ($\delta^{56}Fe$= 0 to +0.5‰) are quite different, providing a means for distinguishing the two processes (see the figure) (*16*, *17*). The distinct $\delta^{56}Fe$ values that are measured for aqueous Fe^{2+} in pore fluids from modern marine sediments that are undergoing sulfide- or DIR-related diagenesis are also different from the iron isotope compositions of hydrothermal fluids ($\delta^{56}Fe$ = –0.5‰) (see the figure), suggesting that marine diagenesis may produce iron isotope variations that are decoupled from the ambient seawater compositions, even in an Archean iron-rich ocean. In the case of Late Archean and Early Proterozoic banded iron formations, the range in $\delta^{56}Fe$ values of magnetite layers has been interpreted to reflect a mixture of iron sources, including hydrothermal fluids and aqueous Fe^{2+} produced by DIR (*18*). It therefore remains unclear if the low-$\delta^{56}Fe$ values measured for diagenetic sulfides in Archean and Proterozoic rocks can be directly related to changes in contents of atmospheric oxygen, as recently proposed (*3*).

Exploring the environments and processes involved in redox cycling of iron is likely to remain a focus of iron isotope geochemistry in the near future. We are just beginning to understand the distinct fractionations that may be associated with specific iron cycling pathways and their overall contributions to the isotopic mass-balance for the iron budget on Earth. Laboratory determinations of isotopic fractionation factors for aqueous species and minerals that occur in nature are critical to interpreting data for natural samples. Thus, experimental work remains a high-priority focus of research in this discipline. In addition, theoretical calculations of isotopic fractionation factors is important for providing a mechanistic basis for interpreting natural isotopic variations, as well as exploring systems that may be difficult to examine in the laboratory (*19*, *20*). This new field of geochemistry is developing rapidly. It seems likely that debate and discussion will continue, reflecting the importance of the issues that are being addressed, such as the origin and evolution of life and changes in the habitable zone on Earth over the past 4 billion years.

References and Notes

1. C. Johnson, B. Beard, F. Albarède, Eds, *Reviews in Mineralogy and Geochemistry*, vol. 55, *Geochemistry of Non-Traditional Stable Isotopes* (Mineralogical Society of America, Washington, DC, 2004).
2. B. L. Beard *et al.*, *Science* **285**, 1889 (1999).
3. O.J. Rouxel, A. Bekker, K.J. Edwards, *Science* **307**, 1088 (2005).
4. B. Beard, C. Johnson, *Rev. Mineral. Geochem.* **55**, 319 (2004).
5. B. L. Beard, C. M. Johnson, *Geochim. Cosmochim. Acta* **68**, 4727 (2004).
6. S. Welch *et al.*, *Geochim. Cosmochim. Acta* **67**, 4231 (2003).
7. S. L. Brantley *et al.*, *Geochim. Cosmochim. Acta* **68**, 3189 (2004).
8. M. S. Fantle, D. J. DePaolo, *Earth Planet. Sci. Lett.* **228**, 547 (2004).
9. K. Yamaguchi *et al.*, *Chem. Geol.* **218**, 135 (2005).
10. S. Levasseur *et al.*, *Earth Planet. Sci. Lett.* **224**, 91 (2004).
11. L. Kump, *Science* **307**, 1058 (2005).
12. C. M. Johnson *et al.*, *Contrib. Mineral. Petrol.* **144**, 523 (2003).
13. T. D. Bullen *et al.*, Geology **29**, 699 (2001).
14. L.R. Croal *et al.*, *Geochim. Cosmochim. Acta* **68**, 1227 (2004).
15. J. Skulan, B. Beard, C. Johnson, *Geochim. Cosmochim. Acta* **66**, 2995 (2002).
16. I. B. Butler *et al.*, *Earth Planet. Sci. Lett.* **236**, 430 (2005).
17. S. Severmann *et al.*, *Eos Trans. AGU* **84**, OS31L-09 (2003).
18. C. M. Johnson *et al.*, *Geochim. Cosmochim. Acta* **69**, 963 (2005).
19. E. Schauble, *Rev. Mineral. Geochem.* **55**, 65 (2004).
20. A. D. Anbar, A. A. Jazecki, T. G. Spiro, *Geochim. Cosmochim. Acta* **69**, 825 (2005).
21. Financial support from NASA and NSF is acknowledged. We thank J. O'Neil, K. Nealson, and L. Kump for comments on the manuscript.

10.1126/science.1112552

STRUCTURAL BIOLOGY

Choosing the Crystallization Path Less Traveled

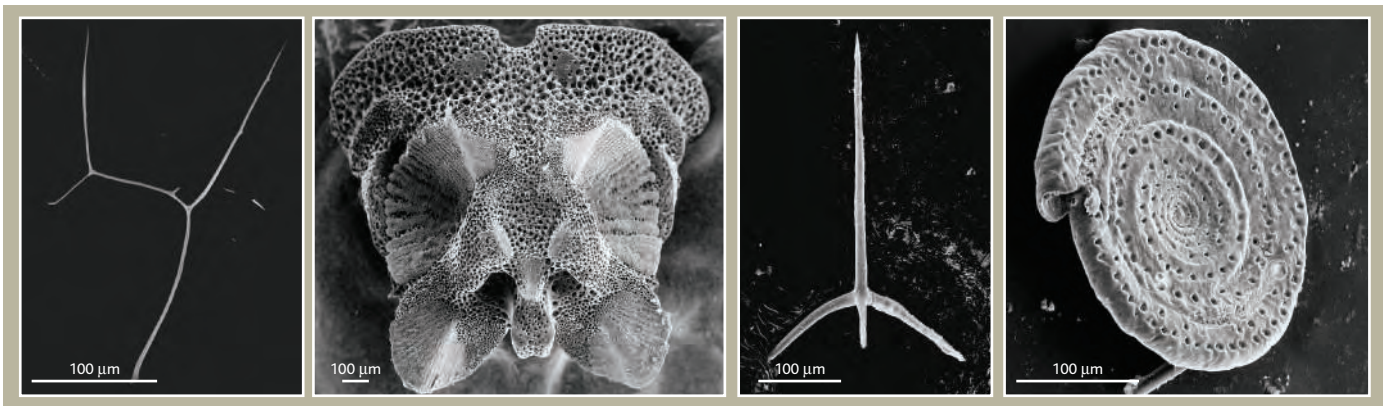
S. Weiner, I. Sagi, L. Addadi

A good way to grow large crystals in desired shapes is by slow cooling of a melt phase of the substance on a small crystal seed. The alternative is to perform the same process with a crystal seed and a slightly supersaturated solution. The crystal seed operates as a nucleating substrate that will reliably induce a crystal of choice to form. Without a nucleating substrate, it is very difficult to control formation of the product. When spontaneous crystallization starts, the process can follow several different pathways that inevitably result in a mass of small crystals of various sizes and shapes. The crystals formed by organisms are usually not at all like that; they are of uniform size and shape and are often aligned. So it would seem that biological systems have opted for the nucleation substrate strategy. And so they do. But biological processes cannot operate at the high temperatures required for a melt phase in inorganic minerals. The supersaturated solution strategy could be viable, but because the solubility of mineral crystals is inevitably low, large solution volumes must be removed to form even relatively small crystals. Amazingly, living organisms have found solutions to both of these problems.

The hallmark of the biological strategy for making certain mineralized skeletal parts is producing the first-formed solid deposits as disordered and often hydrated phases that with time transform into the stable crystalline deposit. The first proven example of such a process in a living organism is from the chiton, a mollusk that has mineralized teeth that are used for scraping rocks to extract algae buried beneath the surface. The outer layer of the tooth contains magnetite, a hard magnetic mineral (*1*). It forms from a disordered ferrihydrite precursor phase (*2*). The inner layer of the tooth contains carbonated apatite, the same mineral present in bone. It forms by way of an amorphous calcium phosphate precursor phase (*3*). In the years following these discoveries, it was shown that ferrihydrite is a precursor phase of magnetite formation in magnetotactic bacteria (*4*). Beniash *et al.* (*5*) later reported that the larva of the sea urchin, an echinoderm, forms its calcitic spicule from an amorphous calcium carbonate precursor phase. Mollusk larvae have also been found to form their aragonitic shells from such a precursor phase (*6*, *7*). And adult sea urchins also apparently follow the same protocol for generating their carbonate skeleton (*8*). Because mollusks and echinoderms are on two different branches of the animal phylogenetic tree, it seems likely that an amorphous calcium carbonate strategy for forming crystals may well be widely used. Indeed, it has been suggested that corals and crustaceans use this approach to produce their skeletons as well (*9*, *10*).

Why should organisms use such a strategy? One reason is that a disordered phase can easily be molded into any shape, whereas a crystal has a strong propensity to adopt a specific shape dictated by the structure of its atomic lattice. There are many examples of beautifully sculpted

The authors are in the Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: steve.weiner@weizmann.ac.il



Larval spicule from the sea urchin *Lytechinus pictus*

Vertebra from the brittle star *Ophiocoma wendtii*

Quadriradiate spicule from the sponge *Kebira uteoides*

Whole shell of the foraminifera *Spirillina* sp.

Skeletons from crystals. Different organisms likely use the same strategy to generate diverse skeletal parts from crystals that arise from a transient amorphous calcium carbonate phase.

carbonate minerals in biology (see the figure). If these are all formed by way of an amorphous calcium carbonate precursor phase, the strategy is most certainly widespread. Smooth curved surfaces in biogenic minerals are often the telltale signs of the involvement of cellular membranes in determining shape.

For about a century it has been known that organisms can produce amorphous calcium carbonate in a stable form. However, amorphous calcium carbonate is inherently unstable. It was only recently synthesized in vitro as a stable phase, and even then under extreme pH conditions (*11*). Specific proteins can induce the formation of stable amorphous calcium carbonate in vitro under nearly physiological pH and temperature conditions (*12*). But stable amorphous calcium carbonate phases differ from each other in that their local atomic organization around the calcium ions varies. Nevertheless, they all have one mole of water per mole of calcium carbonate.

The transient forms of amorphous calcium carbonate that have been observed in organisms are even more enigmatic. They have little or no water, and the nascent short-range order around their calcium ions resembles the calcitic or aragonitic stable phase into which the amorphous phase will transform. All known biogenic amorphous calcium carbonates contain magnesium and phosphate ions (*11*), suggesting critical importance, yet their roles are not clear. In vitro studies of amorphous calcium carbonate formation point to the stabilizing properties of small confined spaces delimited by a membrane or in a hydrophobic medium (*13*). Size itself can be important when nano-sized particles are involved (*14*), as is the case with the spherical subunits that make up amorphous calcium carbonate. Isolating nano-sized amorphous calcium carbonate particles from the external medium until crystallization is desired may well be part of the biological strategy. It may stabilize the transient amorphous phase by preventing contact with water and by preventing contact of the nanoparticles with nucleating substrates.

Crystallization by way of a transient precursor amorphous phase does not preclude the involvement of a designed nucleation substrate. The two processes are clearly integrated in the formation of the sea urchin larval spicule (see the figure). The first step is the oriented nucleation of a single crystal of calcite at a specific location within a large vesicle delimited by a membrane (called a syncitium). The vesicle is then loaded with amorphous calcium carbonate, and a single crystal continues to grow at the expense of the amorphous phase until the entire spicule becomes a single crystal of calcite (*11*). It is conceivable that fusion of membrane-coated amorphous calcium carbonate particles with the large vesicle membrane delivers them uncoated into the vesicle cavity, thus initiating their transformation into crystals. In mollusk shells, the function of a hydrophobic silk fibroin gel phase may be to prevent crystallization until the particles contact the nucleating substrate, ensuring that oriented crystals will form only at the correct location.

We thus come to the peculiar conclusion that organisms achieve their aim of building large single crystals of desired shape by slowly growing them on a small crystal seed or a nucleation substrate by way of a disordered phase. This disordered phase has characteristics similar to a melt phase but does not require high temperatures. This knowledge may well prove useful for fabricating complex-shaped crystalline synthetic materials; because no water is present in the precursor phase, these materials may also have low porosity and improved mechanical properties.

Is it possible that vertebrates also use the transient precursor phase strategy for forming bones and teeth? This option was discussed in the 1970s after it was discovered that in vitro, under nearly physiological conditions, carbonated apatite (the mineral in bone) forms from an amorphous calcium phosphate precursor phase. After it was shown that this phase is not present in mature bone, the possibility of carbonated apatite formation by way of transient precursor phases was also rejected [reviewed in (*15*)]. It may be time to reconsider.

References

1. H. A. Lowenstam, *Geol. Soc. Am. Bull.* **73**, 435 (1962).
2. K. M. Towe, H. A. Lowenstam, *J. Ultrastruct. Res.* **17**, 1 (1967).
3. H. A. Lowenstam, S. Weiner, *Science* **227**, 51 (1985).
4. S. Mann, in *Magnetite Biomineralization and Magnetoreception in Organisms*, J. L. Kirschvink, D. S. Jones, B. J. MacFadden, Eds. (Plenum, New York, 1985), pp. 311–332.
5. E. Beniash, J. Aizenberg, L. Addadi, S. Weiner, *Proc. R. Soc. London Ser. B* **264**, 461 (1997).
6. J. C. Marxen, W. Becker, D. Finke, B. Hasse, M. Epple, *J. Molluscan Stud.* **69**, 113 (2003).
7. I. M. Weiss, N. Tuross, L. Addadi, S. Weiner, *J. Exp. Zool.* **293**, 478 (2002).
8. Y. Politi, T. Arad, E. Klein, S. Weiner, L. Addadi, *Science* **306**, 1161 (2004).
9. A. Meibom *et al.*, *Geophys. Res. Lett.* **31**, L23306 (2004).
10. R. Dillaman, S. Hequembourg, M. Gay, *J. Morphol.* **263**, 356 (2005).
11. L. Addadi, S. Raz, S. Weiner, *Adv. Mater.* **15**, 959 (2003).
12. J. Aizenberg, G. Lambert, L. Addadi, S. Weiner, *Adv. Mater.* **8**, 222 (1996).
13. E. Loste, R. J. Park, J. Warren, F. C. Meldrum, *Adv. Funct. Mater.* **14**, 1212 (2004).
14. M. Li, S. Mann, *Adv. Funct. Mater.* **12**, 773 (2002).
15. M. J. Glimcher, *Philos. Trans. R. Soc. London Ser. B* **304**, 479 (1984).

10.1126/science.1114920

CREDIT: S. WEINER *ET AL.*

INTRODUCTION

Resiliency in the Face of Disaster

The devastating December 2004 Sumatran earthquake and tsunami, the largest modern natural disaster in terms of people and area affected, is just one grim reminder that societies today are facing increasingly diverse and costly natural and human-triggered threats. Many trends are exacerbating the risks. More people are concentrating in coastal areas, where threats of flooding and storms are heightened, and climate change and sea level rise will amplify these risks. Terrorists have attacked the infrastructures of cities. Population movements, along with global trade and transport, heighten the odds of disease pandemics (see the Editorial on p. 989).

Such diverse and, in many cases, unpredictable threats have led to renewed efforts to improve the resiliency of cities and societies overall. This special issue surveys some of these emerging approaches both for preparing for diverse disasters before they happen and for dealing with them afterward.

CREDIT: KIERAN DOHERTY/REUTERS

Two Viewpoints emphasize related aspects of developing social systems that are resilient to unexpected and diverse threats. Allenby and Fink (p. 1034) discuss strengthening infrastructure, cities, and individual businesses; whereas Adger *et al*. (p. 1036) highlight steps for improving the safety of coastal regions. Both emphasize that there are steps that can be taken now that will increase security and help society even if a disaster does not strike. For example, preserving natural ecosystems such as reefs diversifies coastal economies, which can enhance recovery from disasters.

The insurance industry is being called on to help in disaster recovery and to guide preparedness in developed and developing nations alike. Mills (p. 1040) examines how the industry must adapt to deal with climate change. In turn, Linnerooth-Bayer *et al*. (p. 1044) survey some initial approaches that are being developed and tested to expand insurance coverage of these risks to poorer nations.

A major difficulty in recovery from disasters is dealing with the lasting effects they have on the mental health of affected populations. A News story by Miller (p. 1030) describes what is perhaps the broadest effort to date to deal with the psychosocial consequences of disasters and the difficulties of applying Western concepts of mental health to different cultures. Finally, several other aspects of dealing with disasters are explored on *Science*'s Next Wave (see www.sciencemag.org/sciext/disasters/).

–BROOKS HANSON AND LESLIE ROBERTS

CONTENTS

See also the Editorial on p. 989 and related Science's Next Wave material on p. 983 and at www.sciencemag.org/sciext/disasters

Science

Published by AAAS

Utter devastation. More than 75,000 people in Sri Lanka's Ampara district lost their homes in the tsunami; more than 10,000 died.

CREDIT: KIERAN DOHERTY/REUTERS

NEWS

The Tsunami's Psychological Aftermath

The massive psychosocial relief effort has had its problems, but most survivors of the Indian Ocean disaster have shown remarkable resilience. One positive outcome may be a much-needed increase in mental health services for the region

TIRRUKKOVIL, KALMUNAI, AND COLOMBO, SRI LANKA; CHENNAI AND CUDDALORE, INDIA—Two dusty vans pull up at a camp for displaced tsunami survivors near Tirrukkovil, a small fishing village on the east coast of Sri Lanka. Several men hop out and begin setting up a mobile medical clinic. They grab a table and a few plastic chairs and set them up under a thatched roof supported by metal poles. Next comes the pharmacy: a row of medicine jars on a wood plank in a neighboring tent. As word spreads, a line forms to see Dr. Rajandra, a physician sent from the ministry of health.

The first two patients have respiratory infections, which flourish in the close confines of the camp. Next, a young woman in a batik dress complains of a headache that won't go away. She also has abdominal pain, numbness in her extremities, and, when asked, confesses that she hasn't been sleeping well and hasn't felt much like eating. She slumps in her chair and stares blankly at the table but patiently answers the doctor's questions. Her symptoms began about 6 months ago, she says, shortly after the tsunami swept away her house and all her belongings and took the lives of five members of her extended family. Rajandra suspects she's suffering from mild to moderate depression.

The young woman's situation is common, says Boris Budosan, a Croatian-born psychiatrist with the International Medical Corps (IMC) who works in the camps around Tirrukkovil and elsewhere in Ampara district, the region in Sri Lanka hardest hit by the 26 December 2004 tsunami. People in this part of the world almost never speak of mental anguish, Budosan says; it's just not part of the culture. Instead, they complain of various aches, pains, and discomforts that have no apparent physical cause. "In the West you can ask somebody if he's sad," Budosan says. "Here, they don't talk that much about their feelings." Even if the symptoms are masked, Budosan says, he and his colleagues regularly find people living in the camps who suffer from mental problems, especially depression, and who can benefit from counseling or medication.

At the mobile clinic, Rajandra gives the young woman a small plastic bag of anti-anxiety and antidepressant pills, and a worker from IMC gives her a short mental health survey and will pass her name on to a psychiatrist scheduled to visit the camp the following week. Their work is part of a massive effort under way in villages around the Indian Ocean to gauge, and help alleviate, the psychological toll of the tsunami. Perhaps more than any other disaster in recent history, the tsunami triggered an outpouring of aid for the devastated region—along with unprecedented attention to the mental health of the survivors, many of whom saw their children or other family members carried away by the waves. Since the disaster, hundreds of nongovernmental organizations (NGOs) have sent teams to the region to provide various forms of "psychosocial" aid.

Eight months later, the full impact of the tsunami on the mental health of the survivors remains unknown. The World Health Organization (WHO), among others, has estimated that hundreds of thousands of people could suffer lasting psychological effects. Some early evidence, however, suggests that people may be coping better than expected, aided by the Asian emphasis on strong family and community ties. "Given the scale of the catastrophe, the population has been remarkably resilient," says Harry Minas, a psychiatrist at the University of Melbourne in Australia, who has worked with WHO and the Indonesian government on the mental health of tsunami survivors.

Similarly, the verdict is not yet in on how effective the myriad interventions have been. The mostly Western relief groups arrived with abundant good intentions and a wide variety of strategies. But the field of disaster mental health is relatively new, and little

Published by AAAS

research exists on what interventions best stave off long-term psychological problems. Problems have arisen in the aid effort, but many experts say the spotlight on mental health has benefited tsunami survivors and provided political leverage for revamping health policy in the region to include mental health care. This would be a welcome development for a part of the world where mental health problems are thought to take a heavy toll but—as in much of the developing world—are largely unrecognized.

Gauging the impact

One reason solid epidemiological data aren't yet available is that many research teams felt it would be unethical to conduct studies in the immediate aftermath of the disaster. "When we saw what the situation was, we threw our [survey] sheets away," says Prathap Tharyan, a psychiatrist at Christian Medical College in Vellore, India, describing a visit he and colleagues paid to coastal Tamil Nadu state shortly after the tsunami. Instead of doing research, the team found itself on cleanup duty. "Now, 8 months in, we will get approval from our ethics committee and go do a survey," he says. Researchers in other tsunami-affected regions have similar plans.

Some of the most-cited estimates of the toll come from WHO, which in February suggested that up to half of the 5 million people affected by the tsunami would experience moderate to severe psychological distress that would fade without intervention over the course of a year or more. Roughly 5% to 10% would develop more persistent problems, such as depression, posttraumatic stress disorder (PTSD), or other anxiety disorders that would be unlikely to resolve themselves without intervention. And perhaps 1% to 2% would be left with incapacitating mental problems such as major depression or psychosis. WHO also cautioned that the tsunami could trigger acute episodes for thousands of patients with preexisting conditions, especially those who were displaced from psychiatric facilities or lost their medicine in the disaster.

"Talking to psychiatrists [in the affected areas], we get a feeling that our general assessment … was more or less valid," says Shekhar Saxena, WHO's coordinator for mental health evidence and research in Geneva, Switzerland. Preliminary surveys and anecdotal reports suggest that the tsunami has indeed affected people deeply. Many women have been wracked with guilt and anxiety over children they were unable to save. Children have been afraid to leave their parents to go to school. Men have found it hard to return to the sea to fish, and many have turned to alcohol to help cope. Survivors complain of nightmares, flashbacks, and intrusive thoughts of the disaster. One of the challenges for the upcoming epidemiology studies will be to distinguish normal stress and grief responses from psychopathology.

At the same time, many people who have worked with tsunami survivors are struck by their resilience. Asian culture, with its emphasis on group welfare over individual self-reliance, seems to have been a powerful, positive influence. "People came together to support each other and look after the necessities," says Athula Sumathipala, a Sri Lankan psychiatrist who has worked with tsunami survivors in the south and west of the country. "A man who lost his own son would care for someone else's son."

"In India, most people try to deal with grief in the context of community activities," explains Tharyan. "In the West, public grieving is not encouraged, but at funerals here people cry, scream, shout." That has helped people cope, in Tharyan's view. So has religion. "Hinduism has many rituals regarding death," he says. "By the end of the first year, you've done so many rituals you're not grieving anymore."

Sadly, a long-standing familiarity with upheaval and tragedy may also have bolstered the coping mechanisms of many tsunami survivors. "The idea that people who have chronic stress and now have an acute stress will break down is not entirely true," says Saxena. "Sometimes they cope better." In Sri Lanka, ethnic conflict between the majority Sinhalese government and armed Tamil rebels killed at least 60,000 people and displaced at least 800,000 between the early 1980s and a 2002 cease-fire agreement. Much of the violence took place in the Tamil homeland in the north and east of the country, the areas worst hit by the tsunami. Struggle is accepted by people here as a given in life, says P. M. Vincentine, a counselor for SHADE, a psychosocial NGO founded to help war trauma survivors in northern and eastern Sri Lanka but which has shifted its focus to tsunami survivors. "There is a Tamil saying: 'Through the struggle, you have to live,' " he says.

Supportive community. Even in displacement camps, like this one near Tirrukkovil, strong social networks seem to be helping tsunami survivors cope.

Similar sentiments can be heard across the region, from war-torn Aceh province in Indonesia to coastal India. "People here have a tough life to begin with," says Tharyan. "The expectations in life are very different from those in the West." Many people in Tamil Nadu view the tsunami more as the latest obstacle life has thrown at them than as a cataclysmic blow, Tharyan says.

Chaos on the ground

Into this cultural milieu came the psychosocial NGOs. Prominent international relief groups like Médecins Sans Frontières and the Red Cross brought psychiatrists, psychologists, and other workers with extensive field experience in disaster areas. The Scientologists brought "Volunteer Ministers" who trained local people to do "touch assists," a

Talking it out. NGOs have provided a range of counseling services for tsunami survivors.

CREDITS (TOP TO BOTTOM): G. MILLER/*SCIENCE*; PALLAVA BAGLA

Published by AAAS

technique that reputedly eases suffering by restoring communication between injured body parts. Other groups brought everything from trauma counselors to swim instructors to teddy bears.

The upshot, especially in the early days, was chaos. It was nearly impossible to keep track of who was coming and what they were doing, Saxena says. Coordinating the NGOs—a task taken on by WHO and other U.N. agencies in consultation with local governments—has been a major challenge throughout the region, but especially so in Sri Lanka, which has attracted more psychosocial relief groups than other countries. "At times there were more tents set up for the people trying to help than for the people being helped," says Saxena.

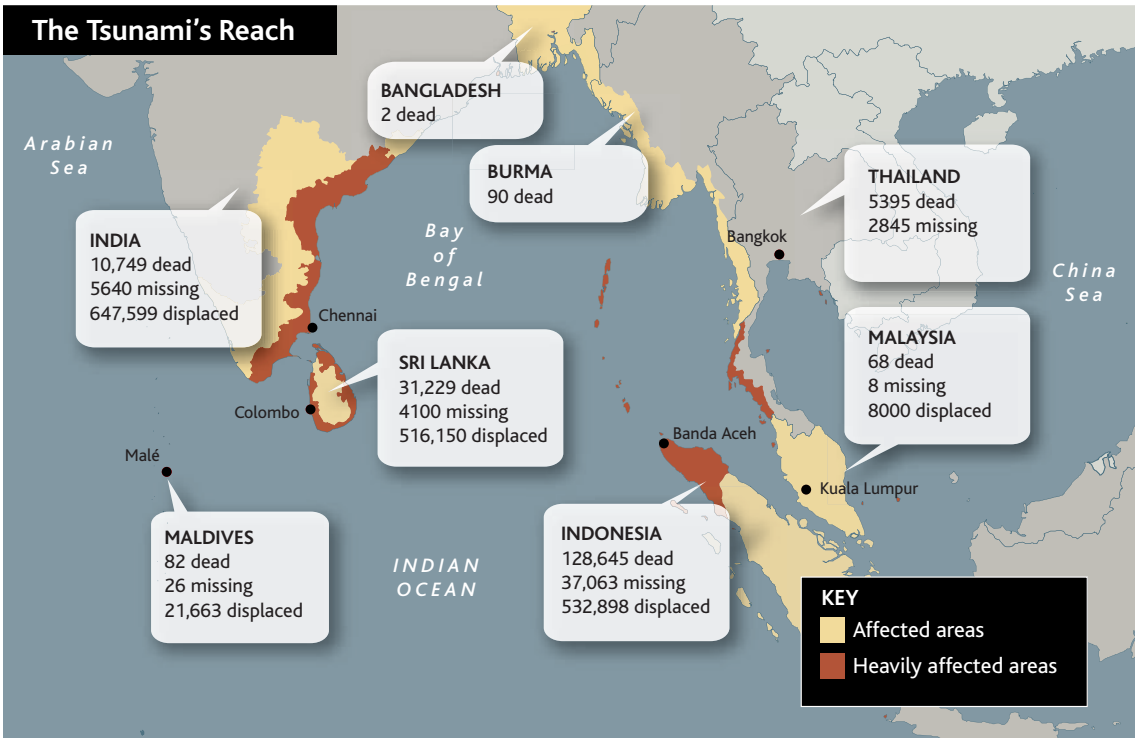
This has occasionally led to friction and even competition as psychosocial NGOs have tried to stake their claims in the refugee camps. At a recent psychosocial coordination meeting in Colombo, several participants said this is still going on. "Kids get attached to volunteers, and then new groups come and offer incentives" for the children to join their activities instead, said T. Gadambanthan, a psychiatrist in the eastern town of Trincomalee. "Children are torn between these loyalties, and it can be traumatic."

Another problem early on was that many aid workers lacked fluency in local languages and knowledge of local culture, says Sumathipala. In the first weeks after the tsunami, for instance, some foreign groups buried bodies in mass graves due to fears of disease. "We have thousands of years of culture here, particularly with regard to death and mourning," says Sumathipala, who believes that casting aside these traditions added to people's suffering. Efforts to dispose of bodies quickly were likely misguided anyway: Research has shown that dead bodies do not pose an imminent threat of disease, and WHO and other groups have discouraged mass burials to allow time for traditional practices.

In terms of psychological services, one of the biggest problems, say many experts, has been an overemphasis on finding and treating cases of PTSD, which is characterized by flashbacks, emotional detachment, sleep difficulties, and other disruptions. Recent years have seen a lively debate among mental health experts over the importance of PTSD in disaster mental health, especially in non-Western settings. The harshest critics see PTSD as a bogus diagnosis—a medicalization of normal grief. But even more moderate experts think the diagnosis has been overemphasized. "It's not at all clear that this is the disorder that burdens people most," says Mark van Ommeren, a specialist in disaster mental health at WHO. "It's only one of many problems that arise after a disaster."

People who come looking for PTSD will find it—and miss half the people who need help, says Minas of the University of Melbourne. In Bosnia, for instance, special centers set up to find and treat people with PTSD "proved to be a disaster," he says. People with war-related depressive and anxiety disorders other than PTSD were overlooked, as were people with preexisting conditions that were exacerbated by the war.

By putting out bulletins and using its contacts with NGOs in the field, WHO has tried, with mixed success, to discourage teams from focusing exclusively on PTSD. The organization has also tried to discourage the use of "single-session debriefing," a controversial intervention intended to reduce posttraumatic stress. In these sessions, survivors are encouraged to relive the traumatic event shortly afterward. But the bulk of research has failed to find evidence that it reduces the incidence of PTSD or other psychological problems, and some studies have suggested that it may even increase the likelihood of problems.



SOURCE: U.N./WFP/ODAP AND USAID

Exporting trauma?

Some critics question the entire enterprise of psychosocial aid to disaster victims, particularly in non-Western countries. Derek Summerfield, a psychiatrist at Maudsley Hospital in London, is among the most vocal. The idea of disaster mental health is "culturally alien" outside the West, Summerfield says: "We can't imagine something like this happening in our countries without our needing counseling, so we take it all to Sri Lanka."

Instead of putting tsunami survivors in our shoes, we should begin by asking what they actually want, says Summerfield, who has worked in Bosnia and other war zones and advised Oxfam and other NGOs on their psychosocial aid programs. In his experience, people want help rebuilding their homes, reestablishing their livelihoods, and getting the children back to school. "They don't want foreigners coming over and saying, 'You've suffered a deep wound in your psyche, and you're going to need our help getting over it,' " he says. R. Thara, a psychiatrist and director of SCARF, a mental health NGO based in Chennai, India, that has done psychosocial work with tsunami survivors in Tamil Nadu, agrees that counseling has been overemphasized. "We had a needs assessment where we asked people what they needed, [and] counseling was the last thing they checked, and probably only because we'd mentioned it." Sumathipala says the same is true in Sri Lanka. "People want material help, people are not asking for counseling," he notes. Two surveys in the north and east of the

country in recent years found that people displaced by the civil war who sought mental health treatment were actually more concerned with finding employment than relief from psychological symptoms such as flashbacks.

Still, Thara, Sumathipala, and others insist that for a small proportion of tsunami survivors, counseling is in fact needed. "As a psychiatrist I believe a certain amount of people would need psychological intervention, including counseling and medication," says Sumathipala.

The critique by Summerfield and others highlights how little is known about the best way to care for the mental health of people who have lived through a disaster. The bulk of the research to date has been carried out in developed countries. As a result, much of what is being tried in Asia, for instance, is based on good intentions rather than good science, says Minas. "I think there's a real obligation to carry out good quality evaluation of what's being done and the consequences of the disaster," he says. "But there's an ethical quandary about what kind of research, and carried out by whom."

In the first few months after the tsunami, says Sumathipala, several foreign researchers were found to be conducting research without approval from any authority in Sri Lanka, or, at least in some cases, approval from their home institution. Sumathipala has been keeping a list of such incidents to push the government to set up a national medical ethics review board. His list includes a Japanese group he says collected blood from tsunami survivors to search for biomarkers of PTSD, without previous approval. A ministry of health official seized their samples and insisted that they get approval, which they subsequently have done. Many groups—university researchers as well as local and foreign NGOs—circulated surveys in the aftermath of the tsunami. Some of these were inappropriate, Sumathipala says, including one distributed by a German group that asked young children detailed questions about sexual abuse.

Minas is working on a set of guidelines that could be used by communities—aided by consultants from local universities and other institutions—to evaluate research proposals in future disasters. Many NGOs are reluctant to do research, Minas says, because they see providing service as the first priority. He thinks that view reflects a misunderstanding of the purpose of research. "Lots of people are ready to just get in and do things without any evidence of whether what they're doing helps people … or maybe even does harm. I think it's negligent to do that without evaluating what's going on."

Looking forward

Despite the glitches in the relief effort and problems with particular NGOs, most observers say the psychosocial response has been beneficial overall. There are also encouraging signs that the influx of money and the expertise of the better trained groups may help pave the way for a stronger mental health care infrastructure.

The need is great. In Sri Lanka alone, WHO estimates that before the disaster some 384,000 people suffered from serious mental disorders such as major depression, bipolar illness, and schizophrenia, and perhaps 2 million who were afflicted by less severe mental disorders. These numbers dwarf even the worst case estimates for mental health problems related to the tsunami. (WHO estimates that 22,000 to 44,000 tsunami survivors in Sri Lanka will develop psychological problems serious enough to require long-term treatment.)

But the country has just 41 psychiatrists, including academics. About four times that many Sri Lankan psychiatrists practice in the United Kingdom, says John Mahoney, WHO's point person in Sri Lanka for mental health. There are no psychiatric nurses and only eight psychiatric social workers for the entire country. Nearly all of these meager resources are concentrated in Colombo, where many patients are relegated to outdated government-run asylums. In June, a Colombo newspaper visited the Mulleriyawa women's asylum near the capital and photographed patients tied to the beds with strips of cloth. The hospital, built to house 400 patients, holds hundreds more, some of whom have been there for decades, Mahoney says: "Places like this shouldn't exist."

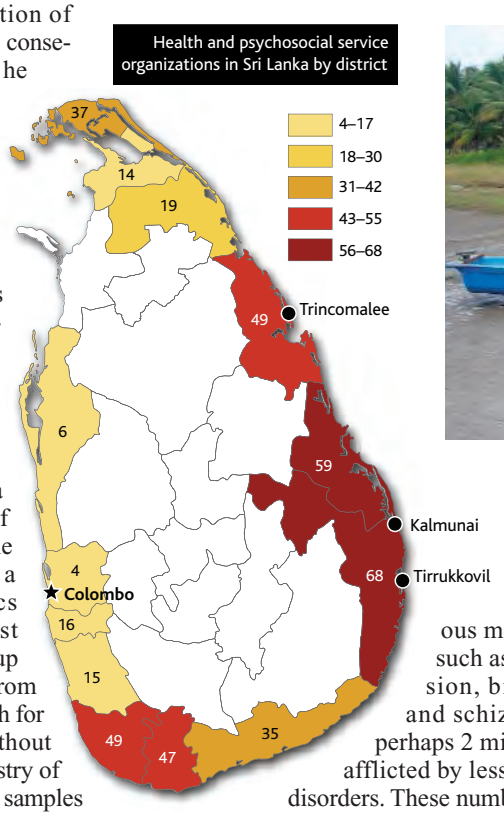
Earlier this summer, the Sri Lankan Ministry of Health approved a plan developed by WHO, in consultation with local health officials, that would close Mulleriyawa and another large mental hospital while vastly increasing access to mental health services, especially for people outside the capital. WHO has pushed for this plan before without much success, but with the momentum from the tsunami behind them, Mahoney is now optimistic that reform is possible. "Now things will happen," he predicts.

A key component of the WHO plan is to provide mental health training to primary care doctors, community health workers, and midwives in Sri Lanka. Rajandra, the ministry of health physician from Tirrukkovil, for example, recently attended a series of mental health workshops put on by IMC. NGOs are providing much of the human—and financial—resources for this training.

Similar plans are in the works for Aceh, where mental health care was virtually nonexistent, and Tamil Nadu in India. The trick, of course, will be to keep the ball rolling after the tsunami aid money dries up. If the plans succeed, however, they may represent the most lasting legacy of the tsunami in terms of mental health.

–GREG MILLER

Reporting for this story was supported by a fellowship from the Carter Center.



Here to help. Many NGOs have been active in Sri Lanka (*left*) and Tamil Nadu, India (*above*), where fishing boats bear the names of the groups that donated them.

CREDITS (LEFT TO RIGHT): HUMANITARIAN INFORMATION CENTRE SRI LANKA; G. MILLER/*SCIENCE*

VIEWPOINT

Toward Inherently Secure and Resilient Societies

Brad Allenby[1]* and Jonathan Fink[2]

Recent years have seen a number of challenges to social stability and order, ranging from terrorist attacks and natural disasters to epidemics such as AIDS and SARS. Such challenges have generated specific policy responses, such as enhanced security at transportation hubs and planned deployment of a global tsunami detection network. However, the range of challenges and the practical impossibility of adequately addressing each in turn argue for adoption of a more comprehensive systems perspective. This should be based on the principle of enhancing social and economic resiliency as well as meeting security and emergency response needs and, to the extent possible, developing and implementing dual-use technologies that offer societal benefits even if anticipated disasters never occur.

Resiliency is defined as the capability of a system to maintain its functions and structure in the face of internal and external change and to degrade gracefully when it must. Developing enhanced resiliency is a rational strategy when the probability and specifics of a particular challenge are difficult to define. However, resiliency is not a global characteristic of a system; it can meaningfully be determined only with reference to an identified system and particular challenges. The Internet, for example, is characterized by a few hubs with high connectivity and an increasing number of other hubs with decreasing connectivity. Such scale-free networks are highly resistant to random failures, in that a substantial number of links can fail and still not affect the performance of the network as a whole. But such architectures are very vulnerable to a deliberate attack directed against the major hubs (*1*). For example, the September 11 attack on the World Trade Center only indirectly affected the Internet, which continued to function almost flawlessly (*2*); it would be much less resilient if directly attacked.

Frequently, a challenge will involve multiple scales, so that overall resiliency requires the ability to understand and take advantage of different initiatives at different levels. For example, designing a building that can be sealed against airborne pathogens is useful, and a number of such buildings in a downtown urban environment will enhance the urban area's overall resiliency against an attack. But designing building-level resilient systems will not substitute for an urban sensor system that enables early and accurate definition of an attack's parameters, nor for the emergency response effort that the city as a whole will need to mount. Analogously, there may be a number of opportunities in the "event life cycle" to implement resiliency strategies. One might invest in avoiding any event in the first place; creating long-term plans that reduce or mitigate threat; generating a warning in time to implement or adjust plans and reduce potential costs; mitigating the event as it occurs; or planning short-term responses and recovery or longer term recovery capabilities.

Some kinds of resiliency are primarily externalities, in that the protection gained provides almost no other benefits, whereas others are dual use and provide substantial economic benefits in addition to resiliency. For example, the communications systems provided to defense and national security organizations are commonly "hardened"; that is, additional technology provides protection against eavesdropping, destructive electromagnetic frequency pulses, and physical intrusion. This extra level of protection obviously adds cost to the system but not additional communications functionality (although the costs are presumably justified by the additional security obtained). In contrast, the creation of internal corporate intranets and support systems for virtual offices and telework capability, which diffuses information assets, can save firms money and make them more resilient against point attacks, as well as natural events such as epidemics (*3*, *4*). More broadly, when a resiliency option is less coupled to other functions, it can be more easily implemented, but it may not offer the additional benefits that strategic investments enhancing resiliency often do.

In general, a portfolio approach based on managing a number of varying risks should be the most efficient. Such an approach should seek to minimize not risk associated with individual events but risk across the social unit as a whole. The portfolio approach is also desirable given the difficulty of unambiguously defining risk and thus investments in resiliency. This ambiguity also serves as an argument for investment in dual-use options where possible—that is, investments that both enhance resilience against attack or disaster and provide additional economic, social, or environmental benefits. Not only are such dual-use technologies important because of resource limitations, but they enhance long-term security as well, for in the longer view a secure society involves innovation in strong infrastructure and social systems as well as in counterterrorism techniques and technologies. Fragile communities are more likely to be susceptible to disaster or attack and to disruption when such events occur and more likely to experience subsequent weakness and failure in the aftermath of an attack.

Network Organization and Urban Systems

Urban systems provide ideal laboratories for understanding resiliency and for developing dual-use technologies, practices, and systems that provide value even if no negative events occur. This is particularly true given accelerating urbanization: Developed countries are already highly urbanized, and the United Nations estimates that the urban populations of Africa, Asia, and Latin America will double over the next 30 years, from 1.9 billion in 2000 to 3.9 billion in 2030. At that point, over 60% of the world's population will live in cities (*5*). Moreover, the cultural, economic, and symbolic importance of urban systems to their societies makes them natural targets for deliberate violence; global transportation networks and high population density make them ideal centers for disease; and the concentration of economic assets and people that characterize them make them highly susceptible to damage from local natural disaster. But cities are not fragile. Indeed, throughout history they have often been destroyed—by fire, by disease, by nuclear attack, by earthquake, and by war—and yet, from 1100 to 1800 only 42 cities worldwide were abandoned after their destruction (*6*).

Cities also present challenging studies in resiliency because their nature is changing rapidly and fundamentally as information becomes an ever more important component of urban structure at all scales (*2*, *7*, *8*). Reliance on information infrastructures by other critical networks, such as transportation, financial, and corporate systems, is also rapidly accelerating (*9*, *10*); cities frequently form critical nodes where these networks intersect and interact. Moreover, the performance characteristics of these networks are also evolving rapidly. In telecommunications, for example, defined and fragile telephone networks have been replaced by Internet-based virtual networks that can be reconfigured and that monitor their own performance and structure and repair themselves in real time (*11*). Similarly, modern computing systems are being designed to con-

[1]Department of Civil and Environmental Engineering, [2]Office of Vice President for Research and Economic Affairs, Arizona State University, Tempe, AZ 85287–7205, USA.

*To whom correspondence should be addressed. E-mail: brad.allenby@asu.edu

tinuously monitor and tune their own performance; adapt to unpredictable conditions (making them resilient and not just engineered for redundancy); predict, prevent, and gracefully recover from failure; and provide safe, secure computing environments (*12*). It is not yet clear how these changes in demographics and information systems will affect the resiliency of urban systems. One immediate result has been increased interest in new tools that aggregate and display complex information patterns at the urban systems level, such as the immersive Decision Theater at Arizona State University (*13*, *14*). Such technologies not only facilitate coordinated emergency management and systemic responses to disasters, but enable better routine management of increasingly complex urban systems and can serve as important educational tools for city managers and the public. They are thus good examples of dual-use technology.

Network-Centric Organizations

The evolution of information-dense urban systems is paralleled by a trend in private firms toward network-centric organizational structures. This parallelism raises a number of questions, including how network-centric firms increase urban system resiliency or, alternatively, vulnerability; whether such firms are indeed more resilient and if so at what scales; and how corporate structure couples community, urban system, regional, and national patterns of social, technological, and economic resiliency. These are highly complex questions requiring further research, but some initial observations can be made.

It is elementary that physical dispersion of assets makes them less subject to point attack or localized disaster such as a tornado or earthquake. A decentralized workforce is also more resilient against a number of other disruptions, including disease (employees who are able to work from home run less risk of infection and help reduce the velocity with which infectious diseases can spread) (*4*). A dispersed workforce enhances resiliency in more subtle ways in addition to the obvious reduction in direct impact. The response to the September 11 attacks indicates that postevent stress and anxiety (the creation of which is a major purpose of many terrorist attacks) can be relieved substantially if arrangements are in place that enable dispersion of the workforce, especially to a home environment where they are both more comfortable and feel themselves less of a potential target (*15*). Ensuring that data and information are not located only in one area, but duplicated in facilities that would not be affected by the same local event, similarly helps protect against catastrophic loss. This was another lesson gained from the September 11 attack on the World Trade Center, where firms such as Lehman Brothers and Cantor Fitzgerald, which had established backup data facilities as part of their business continuity contingency plans, were able to rapidly resume operation (*2*).

But these new patterns of corporate structure have not arisen from concern about terrorism or from seeking resiliency of corporate performance in a risky world. Rather, they reflect economic pressures generated by today's globalized economy with its increasingly dispersed patterns of economic production and increased reliance on information as a critical input to economic activity and production of information as a valuable output (*9*, *16*, *17*). Stronger competition and a more rapidly changing operating environment lead firms and other institutions to adjust in many ways, such as implementing rapid cycle times, learning how to manage and use information networks, developing the ability to absorb and respond to complex information patterns, and emphasizing the knowledge of their workforce as an increasingly critical source of value. Institutional structures are shifting from rigid to more fluid and responsive network-centric organizational patterns, with value and productivity a function of how efficiently the firm can gather and manage knowledge (*2*, *16*).

Concomitantly, the critical infrastructure for many firms is shifting to a substantial degree from their physical assets, such as manufacturing facilities, to knowledge systems and networks and the underlying information and communications technology systems and infrastructure. The functionality that supports corporate, government, and other organizational structures, and most critical corporate data and operational information, now reside on corporate intranets, where they can be accessed from virtually anywhere. This is a costly and potentially disruptive transition in business models, involving substantial changes in many internal organizations such as human resources, real estate management, and information technology management, as well as raising legal, operational, and managerial challenges (*3*). Nonetheless, adoption of these technologies and techniques is driven by competitive pressure, particularly the need to manage costs and increase productivity. Thus, for example, some 30% of the managers at AT&T are completely "virtual" in that they have no assigned office in company-managed buildings, a corporate structure that produces $180 million in business benefits annually, primarily from productivity increases and real estate cost reduction (*4*). Other firms report similar financial benefits (*18*, *19*).

From the perspective of a city, policies that encourage a strong teleworking capability in local firms are ideal dual-use systems: They provide resiliency against disaster or attack, but many important ancillary benefits as well. An urban system with a large number of potential teleworkers can encourage working from home on bad air quality days, or during blizzards or other emergency conditions, or when unanticipated upsets in the traffic networks result in congestion. Moreover, an urban environment that encourages teleworking also provides a higher quality of life; AT&T's data indicate that 81% of its teleworkers name better balance between work and family as a substantial benefit of the practice (*4*). Additionally, some argue that by enabling people to work in their neighborhoods, telework can enhance a sense of community and neighborhood security (*20*).

Developing policies and tools to support implementation of such a dual-use technology is not easy. Novel issues, such as whether a city should invest at the margin in additional transportation infrastructure, such as wider roads, or information infrastructure, such as broadband to the home, are likely to arise. This question is complicated by how investments in information and communication technologies (ICT) interact with the overall evolution of information-dense urban structures. Moreover, the increased reliance on ICT systems and the Internet implied by this process can actually produce vulnerabilities, unless greater emphasis is placed on protecting information infrastructures, especially from deliberate physical or software attack to which they might be most vulnerable given their current structure (*3*). Accordingly, proper network design with hubs geographically separated (and critical ones perhaps duplicated), and network security sensitive to varying degrees of vulnerability of critical network components, including software functionality, should be part of any information and employee dispersion policy or national policy against terrorism.

This point has not been lost on governments. The United States, for example, has issued a series of executive orders and strategies intended to protect ICT infrastructure (*21*, *22*). But vulnerabilities, especially in the private sector, remain widespread, as recent well-reported compromises of consumer and employee data held by major firms indicate (*23*).

At the national scale, the implications of network-centric organizations are profound and only slowly being recognized. For example, reducing unnecessary transportation reduces demand for gasoline and thus enhances energy security. AT&T, for example, estimated that its telework/virtual office program even in 2000 was avoiding some 110 million unnecessary miles of driving per year, avoiding the consumption of more than 5 million gallons of gasoline (and emission of an estimated 50,000 tons of carbon dioxide) (*2*). It also seems likely that, if properly managed, a network-centric society might well be more equitable, more productive, and therefore perhaps less fragile in the face of challenge. Most obviously, many societies use only a small fraction of the intellectual capital available to them; some marginalize women, or noncitizens, but virtually all have relatively arbitrary ages beyond which they marginalize older workers, and most do not have mechanisms to include disabled workers in their economies. Network-centric

structures enable non–place-based access and temporary working arrangements, and cognitive capability built into network tools can facilitate economic integration of the disabled. This enhances not just the economic performance of society, but the quality of life of individuals involved; virtually all marginalized groups are highly interested in participating in the economy if they can and if the work can be structured to suit their requirements, which is precisely the flexibility the network-centric structure can provide. Thus, for example, seniors in the United States report a high interest in continuing to work flexibly (fewer hours, no required office, and no lengthy commutes) (*24*, *25*). On the demand side, the need for adequate knowledge workers will grow substantially as the baby boom generation retires (*25*), and management of pension shortfalls and old-age support policies might well be facilitated by the operational and social flexibility enabled by network-centric economic organization.

The range of ancillary effects discussed in this brief example illustrates the complexities and challenges of adopting the principle of resiliency as a policy and planning touchstone, as well as the potential value of dual-use tools and technologies. Understanding the interplay of these systems and how various investments and policy choices integrated into a resiliency portfolio can simultaneously enhance both security and economic and social stability and growth is not a trivial challenge, but the potential benefits argue strongly for such a course.

References

1. A. Barabasi, *Linked: The New Science of Networks* (Perseus Publishing, Cambridge, MA, 2002).
2. M. Moss, A. Townsend, in *Digital Infrastructures: Enabling Civil and Environmental Systems Through Information Technology*, R. Zimmerman, T. Horan, Eds. (Routledge, London, 2004), pages 141–152.
3. B. R. Allenby, J. Roitz, *Implementing the Knowledge Economy: The Theory and Practice of Telework* (Batten Institute, Darden Graduate School of Business, University of Virginia, Charlottesville, VA, 2003).
4. J. Roitz, B. Nanavati, G. Levy, "Lessons learned from the network-centric organization: 2004 AT&T employee telework results" (AT&T Telework White Paper, AT&T, Bedminster, NJ, 2005).
5. National Research Council, *Cities Transformed* (National Academy Press, Washington, DC, 2003).
6. L. J. Vale, T. J. Campanella, Eds., *The Resilient City* (Oxford Univ. Press, Oxford, 2005).
7. R. Zimmerman, T. Horan, Eds., *Digital Infrastructures: Enabling Civil and Environmental Systems Through Information Technology* (Routledge, London, 2004).
8. M. Amin, in *Digital Infrastructures: Enabling Civil and Environmental Systems Through Information Technology*, R. Zimmerman, T. Horan, Eds. (Routledge, London, 2004), pages 116–140.
9. M. Castells, *The Rise of the Network Society* (Blackwell Publishers, Oxford, 2000).
10. National Research Council, *Information Technology in the Service Society* (National Academy Press, Washington, DC, 1994).
11. AT&T Best Practices, Network Continuity Overview (2005); available at www.att.com/ndr/pdf/cpi_5181.pdf.
12. More information about autonomic computing is available at www-03.ibm.com/autonomic.
13. Arizona State University's Decision Theater Web site is available at http://dt.asu.edu/.
14. J. Fink, F. Steiner, C. Redman, N. Grimm, in *Earth Sciences in the Cities*, G. Heiken, R. Fakundiny, J. Sutter, Eds. (American Geophysical Union Special Publication Series 56), pages 413–426.
15. J. Roitz, personal communication.
16. P. Drucker, *Managing in the Next Society* (St. Martins Press, New York, 2002).
17. L. Edvinsson, M. S. Malone, *Intellectual Capital* (HarperCollins Publishers, New York, 1997).
18. Gartner Group, "Workplace transformation: A workplace imperative" (Report number R-11-0910, Gartner Group, New York, 2000).
19. AT&T Point of View: Remote Teleworking (2004); available at www.business.att.com/emea/english/whitepaper/pdf/remote_working_2004.pdf.
20. B. R. Allenby, D. J. Richards, *Environ. Qual. Manage.* **8**, 3 (2005).
21. U.S. White House, Executive Order 13231, Critical infrastructure protection in the information age, released 16 October 2001; available at www.whitehouse.gov/news/releases/2001/10/20011016-12.html.
22. U.S. White House, "The National Strategy to Secure Cyberspace," February 2003; available at www.whitehouse.gov/pcipb/cyberspace_strategy.pdf.
23. "Information security: The leaky corporation," *The Economist*, 25 June 2005, pp. 57–58.
24. AARP, *Staying Ahead of the Curve: The AARP Work and Career Study* (AARP, Washington, DC, 2002).
25. The Conference Board, *Voices of Experience: Mature Workers in the Future Workforce* (The Conference Board, New York, 2002).

10.1126/science.1111534

VIEWPOINT

Social-Ecological Resilience to Coastal Disasters

W. Neil Adger,[1]* Terry P. Hughes,[2] Carl Folke,[3] Stephen R. Carpenter,[4] Johan Rockström[5]

Social and ecological vulnerability to disasters and outcomes of any particular extreme event are influenced by buildup or erosion of resilience both before and after disasters occur. Resilient social-ecological systems incorporate diverse mechanisms for living with, and learning from, change and unexpected shocks. Disaster management requires multilevel governance systems that can enhance the capacity to cope with uncertainty and surprise by mobilizing diverse sources of resilience.

Human populations are concentrated along coasts, and consequently coastal ecosystems are some of the most impacted and altered worldwide. These areas are also sensitive to many hazards and risks, from floods to disease epidemics. Here, we explore how a better understanding of the linkages between ecosystems and human societies can help to reduce vulnerability and enhance resilience of these linked systems in coastal areas. By resilience, we mean the capacity of linked social-ecological systems to absorb recurrent disturbances such as hurricanes or floods so as to retain essential structures, processes, and feedbacks (*1*, *2*). Resilience reflects the degree to which a complex adaptive system is capable of self-organization (versus lack of organization or organization forced by external factors) and the degree to which the system can build capacity for learning and adaptation (*3*, *4*).

Part of this capacity lies in the regenerative ability of ecosystems and their capability in the face of change to continue to deliver resources and ecosystem services that are essential for human livelihoods and societal development. The concept of resilience is a profound shift in traditional perspectives, which attempt to control changes in systems that are assumed to be stable, to a more realistic viewpoint aimed at sustaining and enhancing the capacity of social-ecological systems to adapt to uncertainty and surprise.

Coastal Hazards and Resilience

Natural hazards are an ongoing part of human history, and coping with them is a critical element of how resource use and human settlement have evolved (*5*, *6*). Globally, 1.2 billion people (23% of the world's population) live within 100 km of the coast (*7*), and 50% are likely to do so by 2030. These populations are exposed to specific hazards such as coastal flooding, tsunamis, hurricanes, and transmission of marine-related infectious diseases. For example, today an estimated 10 million people experience coastal flooding each year due to storm surges and landfall typhoons, and 50 million could be at risk by 2080 because of climate change and increasing population densities (*8*). More and more, adaptive responses will be required in coastal zones to cope with a plethora of similar hazards arising as a result of global environmental change (*9*).

Hazards in coastal areas often become disasters through the erosion of resilience, driven

[1]Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, UK. [2]Centre for Coral Reef Biodiversity, School of Marine Biology and Aquaculture, James Cook University, Townsville QLD 4811, Australia. [3]Centre for Transdisciplinary Environmental Research and Department of Systems Ecology, Stockholm University, SE-10691 Stockholm, Sweden. [4]Center for Limnology, University of Wisconsin, Madison, WI 53706–1492, USA. [5]Stockholm Environment Institute, Box 2142, SE 103 14 Stockholm, Sweden.

*To whom correspondence should be addressed. E-mail: n.adger@uea.ac.uk

by environmental change and by human action (*10*–*12*). For example, when Hurricane Andrew, a powerful category 5 storm, struck Florida in 1992, it caused devastation valued at $26.5 billion and 23 people lost their lives. An equivalent tropical typhoon that ravaged Bangladesh in 1991 resulted in over 100,000 deaths and the displacement of millions of individuals (*13*) from widespread flooding. In Florida, social resilience from strong institutions, early warning systems, and a high capacity to deal with the crisis confined the impact to manageable proportions, whereas social vulnerability in affected areas of Bangladesh caused a human disaster of a far greater scale. Yet adaptive capacity can be increased through purposeful action. Consequently, Bangladesh has reduced mortality associated with typhoons and flooding in the past decade through careful planning focused on the most vulnerable sectors of society (*14*, *15*).

The resilience (or conversely, the vulnerability) of coastal societies is more tightly linked to larger-scale processes today than in the past. For example, economic linkages and the globalization of trade in commodities and ecological goods and services tie regions much more closely together than before (*16*–*18*). In coastal regions, this is often evident in the vulnerabilities created by global tourism (an ecosystem service), where the growing demands of visitors impact previously undeveloped coastal areas (*19*). Similarly, increased mobility of people has spread infectious diseases such as human immunodeficiency virus–acquired immune deficiency syndrome [which have high prevalence in some coastal fishing communities (*20*)], whereas global-scale environmental change is certain to exacerbate vulnerability to vector-borne diseases [e.g., malaria and cholera (*21*, *22*)]. Conversely, greater mobility, improved communications and awareness, and the growth of national and international NGOs that link societies can all strengthen resilience to crises and improve responses when they occur.

During periods of gradual or incremental change, many important sources of resilience may be unrecognized or dismissed as inefficient or irrelevant. Typically, therefore, components of resilience are allowed to decline or are deliberately eliminated because their importance is not appreciated until a crisis occurs. For example, chronic overfishing and declining water quality around coral reefs have made them more vulnerable to cyclones and global warming (*23*). Instead of absorbing recurrent disturbances as they have done for millennia, many overfished and polluted reefs have recently undergone radical regime shifts, where coral populations fail to rebuild after external shocks and have instead been replaced by fleshy seaweeds (*24*, *25*). Rebuilding resilience, by improving water quality and maintaining adequate stocks of herbivores, can promote the regenerative capacity of corals after recurrent disturbances. Thus, loss of ecological and social resilience is often cryptic, and resilience can be eroded or bolstered accidentally or deliberately through human action (*26*).

Resilient social-ecological systems incorporate diverse mechanisms for coping with change and crisis (*27*, *28*). In ecosystems, biodiversity, functional redundancy, and spatial pattern can all influence resilience. Biodiversity enhances resilience if species or functional groups respond differently to environmental fluctuations, so that declines in one group are compensated by increases in another (*24*, *29*). Spatial heterogeneity can also confer resilience, as when refuge areas provide sources of colonists to repopulate disturbed regions (*30*). Similarly, in social systems, governance and management frameworks can spread risk by diversifying patterns of resource use and by encouraging alternate activities and lifestyles. Such practices sustain ecosystem services, analogous to the way that management of a diverse portfolio sustains the growth of investments in financial markets (*31*).

After catastrophic change, remnants ("memory") of the former system become growth points for renewal and reorganization of the social-ecological system (*28*). Ecological memory is conferred by biological legacies that persist after disturbance, including mobile species and propagules that colonize and reorganize disturbed sites and refuges that support such legacies and mobile links (*30*, *32*). Social memory comes from the diversity of individuals and institutions that draw on reservoirs of practices, knowledge, values, and worldviews and is crucial for preparing the system for change, building resilience, and for coping with surprises (*33*).

Fig. 1. Mosque and crop field in Banda Aceh, Indonesia, before (**top**) and after (**bottom**) the 2004 Southeast Asia tsunami, illustrating the impact of natural disasters on the delivery of ecological goods (agriculture) and the social cohesion of resilient societies.

Responding to Change in Coastal Areas

How can coastal zones be transformed into systems that are more resilient and adaptive to a rising incidence of large disturbances? We review two case studies as examples. The first is the 2004 Asian tsunami, which shows that social-ecological resilience is an important determinant of both the impacts of the tsunami and of the reorganization by communities after the event. The second is from research on planning for and adaptating to severe storms and climate change in coastal zones and on small islands. In both cases, individuals and communities undertake adaptive strategies that involve the mobilization of assets, networks, and social capital both to anticipate and to react to potential disasters. Crucially, the causes of vulnerability are embedded in the political economy of resource use and the resilience of the ecosystems on which livelihoods depend.

The 2004 Asian tsunami. On 26 December 2004, countries in South and Southeast Asia experienced an enormous tsunami associated with the second-largest earthquake in the instrumental record. Coastal areas in parts of Indonesia, Thailand, and Malaysia closest to the epicenter received little or no warning (Fig. 1). A key lesson is that resilient social-ecological systems reduced vulnerability to the impacts of the tsunami and encouraged a rapid, positive response. This response needs to be sustained in the longer term, long after the tsunami fades from global news reports.

Chronic degradation of local environments has influenced the short- to medium-term impact of the tsunami and will continue to shape the longer-term options for rebuilding. In Banda Aceh, Indonesia, the presence or absence of sand dunes, mangrove forests, and coral reefs made no dif-

ference in the impact of giant waves that penetrated kilometers inland. Further from the epicenter, however, in Sri Lanka, the energy of smaller waves was reduced by natural barriers (*34*, *35*). Moreover, wherever ecosystems have been undermined, the ability to adapt and regenerate has been severely eroded. For example, throughout coastal Asia, deforestation of mangrove for intensive shrimp farming, a lucrative export industry, has reduced the livelihood options available to local farming and fishing communities (*36*). In many locations, environmental degradation such as land clearing, coastal erosion, overfishing, and coral mining has reduced the potential for economic recovery from the tsunami because of the loss of traditional income sources related to coastal ecosystems rich in biodiversity and ecological functions.

Social resilience, including institutions for collective action, robust governance systems, and a diversity of livelihood choices are important assets for buffering the effects of extreme natural hazards and promoting social reorganization (Table 1). Coastal communities harboring knowledgeable, prepared, and responsive institutions are more likely to be able to prevent the tsunami from making the transition from extreme natural hazard to longer-term social disaster (*37*). For instance, fishing communities on Simeulue Island, west of Sumatra and close to the epicenter of the earthquake causing the tsunami, and on Surin Island, Thailand, survived the tsunami thanks to inherited local knowledge of tsunamis and to institutional preparedness for disasters.

There has been a well-meaning rush by organizations and international aid agencies to apply engineering approaches to rebuilding coral reefs damaged by the tsunami by transplanting corals and constructing miniature artificial reefs. However, none of these engineering interventions actually work at meaningful scales or provide realistic solutions to the increased global threats to coral reefs. Fundamentally, the upsurge in investment in artificial rehabilitation of reefs is misguided because it fails to reverse the root causes of regional-scale degradation. Before the tsunami, runoff from land, overfishing, destructive fishing practices (bombing and poisoning), and climate change had already seriously degraded many reefs. Throughout the region, chronic pollution and overfishing of herbivorous fishes have promoted blooms of turfing and fleshy seaweed that overgrow and smother juvenile corals. Regeneration of damaged reefs continues to be impaired by these and other ongoing human impacts. Now the tsunami has added to the destruction in many locations, smashing corals and smothering reefs with choking sediments. Realistically, regeneration processes in the wider seascape are the only means by which coral reefs can reestablish after large-scale damage (*25*). Consequently, restoration efforts should focus on improving water quality and restoring depleted fish stocks to bolster the innate resilience of coral reefs (*24*). Scarce reconstruction aid should not be squandered on simplistic and ineffective reef rehabilitation projects. Rather, support should be directed to provide ecologically sustainable, long-term employment for coastal communities, to eliminate poverty, and to improve local and regional governance systems for managing the natural resilience of coral reefs.

The 2004 Asian tsunami tragedy demonstrates that formal and informal institutions with the capacity to respond to rapid change in environmental and social conditions are a key to mitigating the social effects of extreme natural hazards. Rather than attempting to reduce or eliminate inherent change and variability (the conventional engineering approach to "control" nature), governance systems, from governments through to local marine and land tenure systems, need to focus on sustaining and enhancing the sources of resilience of societies and their life-supporting ecosystems. The hidden success story of the tsunami was the prevention of widespread secondary mortality of injured and traumatized victims from infection and disease, due in large part to the unprecedented scale of national and international responses.

Coping and adapting to hurricanes. Hurricanes, typhoons, and their related impacts affect societies throughout the world. They do so both directly through acute damage on human settlement, often with major loss of life, and indirectly through their impact on coastal ecosystems such as coral reefs, seagrass beds, and mangroves (*38*, *39*) that support local societies and economies. There is growing consensus than human-influenced climate changes are now evident in hurricane regions and are likely to affect hurricane intensity and rainfall (which cause much of the damage), although the effect of climate change on hurricane frequency in the future remains uncertain (*40*). Although the costs of weather and climate events in terms of economic damage and lives at risk are rising through time, the observed increases are caused by changing social vulnerabilities as much as by changing physical hazards (*39*).

In the Caribbean, responses to hurricanes and their effectiveness depend on social and ecological resilience. The Cayman Islands, for example, has implemented adaptation actions at national and community levels, building both preparedness and community resilience. The implementation of these activities followed economic and ecological impacts of three major hurricanes in 1988 (Gilbert), 1998 (Mitch), and 2000 (Michelle). The resilience of the islands was subsequently put to the test by Hurricane Ivan (2004) and was demonstrably improved. Adaptations included changes in the rules and governance of hurricane risk, change in organizations, establishment of early warning systems, and promotion of self-mobilization in civil society and private corporations. Social learning, the diversity of adaptations, and the promotion of strong local social cohesion and mechanisms for collective action have all enhanced resilience and continue to guide planning for future climate change (*41*). After Hurricane Ivan in 2004, private sector interests in tourism and banking accelerated recovery by rebuilding public infrastructure such as roads and electricity supply. In Trinidad and Tobago, networks associated with present-day coral reef management also play a key role in disaster preparedness and in building resilience (*42*, *43*). Hence, social resilience to disasters in the Caribbean has been promoted through a wide diversity of institutional forms.

However, large sections of society in the Caribbean region remain vulnerable, and cur-

Table 1. Examples of local- and regional-scale actions to enhance resilience in social-ecological systems exposed to abrupt change.

Elements of vulnerability	Local action	National and international action
Exposure and sensitivity to hazard	Maintenance and enhancement of ecosystem functions through sustainable use Maintenance of local memory of resource use, learning processes for responding to environmental feedback and social cohesion	Mitigation of human-induced causes of hazard Avoidance of perverse incentives for ecosystem degradation that increase sensitivity to hazards Promotion of early warning networks and structures Enhancement of disaster recovery through appropriate donor response
Adaptive capacity	Diversity in ecological systems Diversity in economic livelihood portfolio Legitimate and inclusive governance structures and social capital	Bridging organizations for integrative responses Horizontal networks in civil society for social learning

rent adaptation processes are not always appropriate or effective. The impacts of Hurricane Mitch on Honduras, Nicaragua, and El Salvador in 1998 were exacerbated by unsound economic policies, such as export-driven agriculture. Farmers who had adopted modern management practices suffered greater losses than those who had more traditional agro-ecological practices (*44*). Industrialized agricultural practices also generated unexpected impacts and risks, such as the release of 70 tons of toxic pesticides into the environment in Honduras after the destruction of several warehouses, exposing rural populations to long-term harm (*45*). Even today, the lessons of implementing postdisaster planning to increase adaptive capacity do not appear to have been learned by many of the states that were impacted by Hurricane Mitch.

In summary, the social-ecological resilience of tsunami- or hurricane- and typhoon-affected regions involves many elements and actions (Table 1), and each of these involves human agency. Exposure to hazards can often be modified through government interventions or informal norms that regulate the use of coastal ecosystems. Reducing the perverse incentives that destroy natural capital and thus exacerbate vulnerability in the first place should, in many cases, be the priority. Networks and institutions that promote resilience to present-day hazards also buffer against future risks, such as those associated with climate change. Effective multilevel governance systems are critical for building capacity to cope with changes in climate, disease outbreaks, hurricanes, global market demands, subsidies, governmental policies, and other large-scale changes. The challenge for social-ecological systems is to enhance the adaptive capacity to deal with disturbance and to build preparedness for living with change and uncertainty (*28*).

Conclusions

The case for building resilience in coastal regions is urgent, given trends in human settlement, resource use, and global environmental change. Two-thirds of the coastal disasters recorded each year are associated with extreme weather events, such as storms and flooding, that are likely to become more pervasive threats because of anthropogenically driven shifts in Earth's climate and sea level rise. These risks in particular are exacerbated by human action, raising the possibility that greenhouse gas emitters may one day become legally liable for impacts (*46*). Clearly, the reduction of greenhouse gas emissions is necessary in this context but not sufficient in the management of hazards in coastal regions. Already, the resilience of many social-ecological systems has been eroded, particularly in vulnerable, marginalized societies.

The capacity of coastal ecosystems to regenerate after disasters and to continue to produce resources and services for human livelihoods can no longer be taken for granted. Rather, socio-ecological resilience must be understood at broader scales and actively managed and nurtured. Incentives for generating ecological knowledge and translating it into information that can be used in governance are essential. Multilevel social networks are crucial for developing social capital and for supporting the legal, political, and financial frameworks that enhance sources of social and ecological resilience (*33*, *47*). The sharing of management authority requires cross-level interactions and cooperation, not merely centralization or decentralization. In many cases, improved, strong leadership and changes of social norms within management organizations are required to implement adaptive governance of coastal social-ecological systems. There is no time to waste.

References and Notes

1. C. S. Holling, *Annu. Rev. Ecol. Syst.* **4**, 1 (1973).
2. B. Walker, C. S. Holling, S. Carpenter, A. Kinzig, *Ecol. Soc.* **9**(no. 2), U165 (2004); available online at www.ecologyandsociety.org/vol9/iss2/art5.
3. S. Carpenter, B. Walker, J. M. Anderies, N. Abel, *Ecosystems* **4**, 765 (2001).
4. C. Folke *et al.*, *Ambio* **31**, 437 (2002).
5. J. Diamond, *Guns, Germs, and Steel* (Norton, New York, 1999).
6. R. C. Sidle *et al.*, *Quat. Int.* **118-119**, 181 (2004).
7. C. Small, R. J. Nicholls, *J. Coast. Res.* **19**, 584 (2003).
8. R. J. Nicholls, *Glob. Environ. Change* **14**, 69 (2004).
9. R. F. McClean, A. Tsyban, in *Climate Change 2001: Impacts, Adaptation, and Vulnerability: IPCC Working Group II*, J. J. McCarthy, O. Canziani, N. A. Leary, D. J. Dokken, K. S. White, Eds. (Cambridge Univ. Press, Cambridge, 2001), pp. 345–379.
10. P. O'Keefe, K. Westgate, B. Wisner, *Nature* **260**, 566 (1976).
11. P. Blaikie, T. Cannon, I. Davis, B. Wisner, *At Risk: Natural Hazards, People's Vulnerability, and Disasters* (Routledge, London, 1994).
12. B. L. Turner II *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8074 (2003).
13. F. Miller, F. Thomalla, J. Rockström, *Sustainable Dev. Update* **1**, 2 (2005).
14. M. M. Q. Mirza, *Clim. Policy* **3**, 233 (2003).
15. S. Huq, Z. Karim, M. Asaduzzaman, F. Mahtab, Eds., *Vulnerability and Adaptation to Climate Change in Bangladesh* (Kluwer, Dordrecht, Netherlands, 1999).
16. K. O'Brien *et al.*, *Glob. Environ. Change* **14**, 303 (2004).
17. W. N. Adger, H. Eakin, A. Winkels, in *Global Environmental Change and the South-East Asian Region: An Assessment of the State of the Science*, L. Lebel, Ed. (Island Press, Washington, DC, in press).
18. W. N. Adger, N. Brooks, in *Natural Disasters and Development in a Globalising World*, M. Pelling, Ed. (Routledge, London, 2003), pp. 19–42.
19. K. Brown, R. K. Turner, H. Hameed, I. Bateman, *Environ. Conserv.* **24**, 316 (1997).
20. E. H. Allison, J. A. Seeley, *Fish Fish.* **5**, 215 (2004).
21. R. R. Colwell, *Science* **274**, 2025 (1996).
22. C. D. Harvell *et al.*, *Science* **285**, 1505 (1999).
23. T. P. Hughes *et al.*, *Science* **301**, 929 (2003).
24. D. Bellwood, T. Hughes, C. Folke, M. Nyström, *Nature* **429**, 827 (2004).
25. T. Hughes *et al.*, *Trends Ecol. Evol.* **20**, 380 (2005).
26. C. Folke *et al.*, *Annu. Rev. Ecol. Evol. Syst.* **35**, 557 (2004).
27. L. Gunderson, C. S. Holling, Eds., *Panarchy: Understanding Transformation in Human and Natural Systems* (Island Press, Washington, DC, 2002).
28. F. Berkes, J. Colding, C. Folke, Eds., *Navigating Social-Ecological Systems: Building Resilience for Complexity and Change* (Cambridge Univ. Press, Cambridge, 2003).
29. A. R. Ives, K. Gross, J. L. Klug, *Science* **286**, 542 (1999).
30. M. Nyström, C. Folke, *Ecosystems* **4**, 406 (2001).
31. R. Costanza *et al.*, *Bioscience* **50**, 149 (2000).
32. T. Elmqvist *et al.*, *Front. Ecol. Environ.* **1**, 488 (2003).
33. C. Folke, T. Hahn, P. Olsson, J. Norberg, *Annu. Rev. Environ. Res.*, in press.
34. P. L.-F. Liu *et al.*, *Science* **308**, 1595 (2005).
35. F. Dahdouh-Guebas *et al.*, *Curr. Biol.* **15**, R443 (2005).
36. W. N. Adger, P. M. Kelly, N. H. Ninh, Eds., *Living with Environmental Change: Social Vulnerability, Adaptation and Resilience in Vietnam* (Routledge, London, 2001).
37. J. Colding, T. Elmqvist, P. Olsson, in *Navigating Social-Ecological Systems: Building Resilience for Complexity and Change*, F. Berkes, J. Colding, C. Folke, Eds. (Cambridge Univ. Press, Cambridge, 2003), pp. 163–185.
38. A. E. Lugo, C. S. Rogers, S. W. Nixon, *Ambio* **29**, 106 (2000).
39. R. A. Pielke Jr. *et al.*, *Nat. Hazards Rev.* **4**, 101 (2003).
40. K. Trenberth, *Science* **308**, 1753 (2005).
41. E. L. Tompkins, *Glob. Environ. Change* **15**, 139 (2005).
42. E. L. Tompkins, W. N. Adger, *Ecol. Soc.* **9**(no. 2), U190 (2004); available online at www.ecologyandsociety.org/vol9/iss2/art10.
43. K. Brown, E. L. Tompkins, W. N. Adger, *Making Waves: Integrating Coastal Conservation and Development* (Earthscan, London, 2002).
44. E. Holt-Gimenez, *Agric. Ecosyst. Environ.* **93**, 87 (2002).
45. K. Jansen, *Dev. Change* **34**, 45 (2003).
46. M. R. Allen, R. Lord, *Nature* **432**, 551 (2004).
47. T. Dietz, E. Ostrom, P. C. Stern, *Science* **302**, 1907 (2003).

48. We thank the Christensen, McDonnell, and Packard foundations for support through the Resilience Alliance. W.N.A. acknowledges the U.K. Natural Environment Research Council, the Engineering and Physical Sciences Research Council, and the Economic and Social Research Council for financial support through the Tyndall Centre. C.F. acknowledges the support of the Swedish Research Council for Environment, Agricultural Sciences, and Spatial Planning. T.P.H. was supported by grants from the Australian Research Council (ARC) and an ARC Federation fellowship.

10.1126/science.1112122

VIEWPOINT

Insurance in a Climate of Change

Evan Mills*

Catastrophe insurance provides peace of mind and financial security. Climate change can have adverse impacts on insurance affordability and availability, potentially slowing the growth of the industry and shifting more of the burden to governments and individuals. Most forms of insurance are vulnerable, including property, liability, health, and life. It is incumbent on insurers, their regulators, and the policy community to develop a better grasp of the physical and business risks. Insurers are well positioned to participate in public-private initiatives to monitor loss trends, improve catastrophe modeling, address the causes of climate change, and prepare for and adapt to the impacts.

Business and science meet in the wake of disasters. The insurance sector is a lightning rod, serving as global integrator of impacts across all sectors of the economy, and messenger of these impacts through the terms and price signals it projects to its customers (*1*). As the world's largest industry [it would be the third largest country if its $3.2 trillion in yearly revenues were compared with national gross domestic products (GDPs)], the implications of rising disaster losses on insurers are as important as defining the industry's role in furthering understanding of the problem and advancing loss-prevention solutions.

The insurance "industry" is non-monolithic, with considerable regional variations in coverages, hazard exposure, and regulation within and among countries. Insurance penetration averages 9% of GDP ($2750/capita) in industrialized countries and 5% of GDP ($25/capita) in developing countries and economies in transition (*2*). Although 12% of premiums today come from this latter market, at current growth rates it will constitute half of the global market within a few decades. Insurance payouts for weather-related disasters in the developing world are today three times the amount provided by international aid (*3*).

Insurance is part of a broader public-private patchwork for spreading risks across time, over large geographical areas, and among diverse social and commercial communities. Not all natural hazards are insured. In some cases (e.g., flood, crop), public and private agencies share the risk. The growing repository of insurance loss data—considered among the best sources of disaster statistics (*4*)—augments geophysical observing systems with trends in economic impacts.

*The author is at the Lawrence Berkeley National Laboratory, MS 90-4000, Berkeley, CA 94720, USA. E-mail: emills@lbl.gov

The availability and affordability of insurance are grist for economic development and the financial cohesion of society, as well as security and peace of mind in a world where the knowledge of hazards lags their evolution. Unanticipated changes in the nature, scale, or location of hazards are among the most important threats to the insurance system. History has shown that society in general, and insurers in particular, are often caught unprepared for ostensibly "inconceivable" disasters. This reflects, in part, the recurring social miscalculation of using the past to predict the future while underinvesting in disaster preparedness. Be it the attacks of "9/11" or Hurricane Andrew, expectations based on past experience led to complacency and dramatic underestimation of exposure. An eye-opening insurance industry report from the mid-1980s (*5*) highlighted the importance of anticipating multiple large events in a single year, yet exposures are still often expressed in terms of probable maximum losses for single events rather than for entire insurance "seasons." The limitations of this approach were evident in the 2004 U.S. hurricane season and its $60 billion in economic losses (of which half were insured).

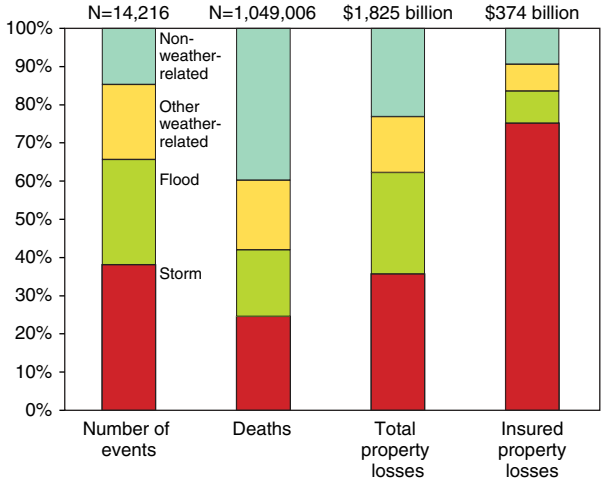


Fig. 1. Global impacts of natural disasters from 1980 to 2004. Insured property losses are dominated by storm events due to risk-selection preferences of insurers and coverage of flood and crop exposures by public entities, and low penetration of earthquake insurance. Economic values are inflation-adjusted to 2004 levels. [Source: Munich Re, NatCatSERVICE]

The weather-dependent share of global insured catastrophe losses (~90%) is greater than that experienced by the economy as a whole (~75%) (Fig. 1). This, coupled with the increase in the number, cost, and variability of such losses (Fig. 2), has brought some insurers, reinsurers, and their trade associations to view climate change as a strategic factor in their future (*6*–*8*).

Virtually all segments of the industry have a degree of vulnerability to the likely impacts of climate change, including those covering damages to property (structures, automobiles, marine vessels, aircraft); crops and livestock; pollution-related liabilities; business interruptions, supply-chain disruptions, or loss of utility service; equipment breakdown arising from extreme temperature events; data loss from power surges or outages; and a spectrum of life and health consequences (*1*).

Specific technical risks include the following: (i) Shortening times between loss events. (ii) Changing absolute and relative variability of losses. (iii) Changing structure of types of events. (iv) Shifting spatial distribution of events. (v) Damage functions that increase exponentially with weather intensity (e.g., wind damages rise with the cube of the speed). (vi) Abrupt or nonlinear changes in losses. (vii) Widespread geographical simultaneity of losses (e.g., from tidal surges arising from a broad die-off of protective coral reefs or disease outbreaks on multiple continents). (viii) More single events with multiple, correlated consequences. This was well evidenced in the pan-European heat catastrophe of 2003—where temperatures were six standard deviations from the norm (*9*). Immediate or delayed impacts included extensive human morbidity and mortality, wildfire, massive crop losses, and the curtailment of electric power plants owing to the high temperature or lack of cooling water. (ix) More hybrid events with multiple consequences [e.g., El Niño–Southern Oscillation (ENSO)–related rain, ice storms, floods, mudslides, droughts, and wildfires].

Specific market-based risks include the following: (i) Historically based premiums that

lag behind actual losses (especially for life insurers, where premiums may be fixed over long periods). (ii) Failing to foresee and keep up with changing customer needs arising from the consequences of climate change. (iii) Unanticipated changes in patterns of claims, and associated difficulty in adjusting pricing and reserve practices to maintain profitability. (iv) Responses of insurance regulators (*10*). (v) Reputational risks falling on insurers who do not, in the eyes of consumers, do enough to prevent losses arising from climate change. (vi) Stresses unrelated to weather but conspiring with climate change impacts to amplify the net adverse impact. These include drawdowns of reserves due to earthquakes or terrorist attacks and increased competition from self-insurance or other competing methods of risk-spreading (especially if relatively low-risk customers shift to those products).

Observed Trends

It is widely recognized that the costs of weather-related natural disasters have been rising. The impacts include an elevated need for assistance from outside impacted areas (*11*) and a shrinking gap between insurance premiums and losses (Fig. 3).

From 1980 through 2004, the global economic costs of such events totaled US$1.4 trillion ($2004), of which $340 billion were insured. To put the burden of these costs on insurers in perspective, recent average annual losses surpass those experienced in the aftermath of the 9/11 attacks in the United States. These costs are substantial for insurers and their customers, leading to industry-wide unprofitability in the worst years (even including investment gains), abrupt price increases, and isolated bankruptcies. Although headline-catching catastrophic events are the most visible, the average aggregate insured cost of smaller events is 60% of the total (*1*).

The insured share of total economic losses from weather-related catastrophes is rising, increasing from a negligible fraction in the 1950s to 25% in the last decade. The ratio has climbed more quickly in the United States, with more than 40% of the total disaster losses insured in the 1990s (*12*).

For several reasons, the cited magnitude of losses systematically underestimates actual costs to insurers and the broader economy because, although large in aggregate, small events are rarely captured in these statistics (especially in the developing world). For example, the Property Claims Services (PCS), which compiles data for U.S. insurers, includes only those events with costs above a threshold of US$25 million. Among the types of events often excluded, power outages in the United States alone are estimated to result in a cost of US$80 billion per year (*13*) [and weather-related events account for 60% of the customers affected by disturbances on the bulk power grid (*14*)], and lightning strikes cause billions of dollars of losses each year, as do damages from soil subsidence (*1*), the melting of permafrost (*15*), and wildfire (*16*). No winter storms were included in the PCS statistics for the 46-year period from 1949 to 1974, and few were included thereafter (*17*). Similarly, aggregate weather-related vehicle accidents are typically not tracked. Furthermore, trends toward increasing deductibles and decreasing insured limits, as well as a strong shift toward self-insurance, lower year-to-year insurance payouts for like events. Finally, the figures cited above largely exclude weather-related life and health costs; restrictions on trade, travel, and tourism; disaster preparedness; and evacuations, energy price increases, and other second-order market impacts of severe weather.

Attribution

Socioeconomic and demographic trends clearly play important—and likely dominant—roles in the observed upward loss trends (*18*). As recognized by insurers and others, migration of populations to flood- and fire-prone areas, increasing reliance on vulnerable electric power grids, and rising material wealth are among the many drivers. Changes in the incidence and impacts of extreme weather events and sea-level rise can also be observed (*19*–*22*).

Global weather-related losses in recent years have been trending upward much faster than population, inflation, or insurance penetration, and faster than non–weather-related events (Fig. 2D). By some estimates, losses have increased by a factor of 2, after accounting for these factors plus increased density of insured values (*23*, *24*). The Association of British Insurers states that changes in weather could already be driving UK property losses up 2 to 4% per year (*7*) owing to increasing extreme weather events. Specific event types have increased far more quickly than the averages. For example, damages from U.S. storms grew 60-fold to US$6 billion/year between the 1950s and the 1990s (*21*).

According to the latest Intergovernmental Panel on Climate Change (IPCC) assessment, climate change has played a role in the rising costs of natural disasters (*1*). As an illustration of the linkages, the distribution and frequency of lightning strikes is expected to shift under climate change (*25*), and insurers indeed observe a notable increase in losses during periods of elevated temperatures (Fig. 4) (*6*).

Many human activities mask losses that would otherwise manifest in the trend data. These include improved building codes, early-warning systems, flood control, electric load-shedding to avoid blackouts during heatwaves, disaster preparedness and response, and land-use planning. Insurer actions to reduce their exposures produce a dampening effect on ob-

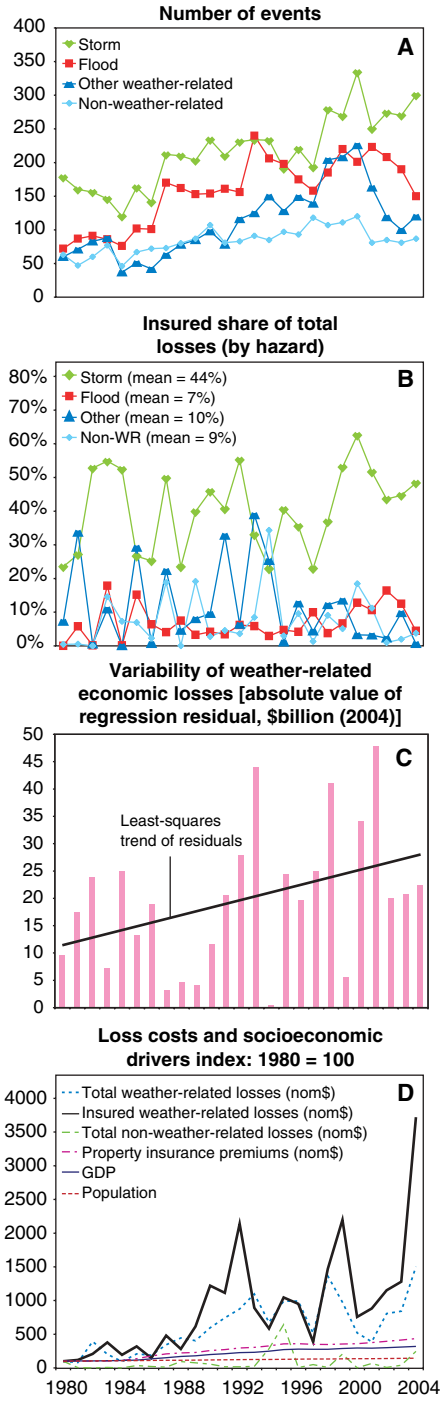


Fig. 2. (**A** to **D**) Trends in (A) global numbers of weather-related events, (B) insured share, (C) variability of total losses, and (D) economic impacts and demographic drivers. Insured and total property losses ($45 billion and $107 billion in 2004, respectively) are rising faster than premiums, population, or economic growth. Data exclude health and life insurance premiums and losses. Non-inflation–adjusted economic data are shown in relation to GDP. Inflation-adjusted economic losses from catastrophic events rose by 8-fold between the 1960s and 1990s and insured losses by 17-fold (*11*). [Sources: Natural hazard statistics and losses from Munich Re, NatCatSERVICE; Premiums from Swiss Re, *Sigma*]

served insured costs. Untangling these offsetting factors is a necessary part of any comprehensive attribution analysis and has not been dealt with satisfactorily in the literature.

In any event, the consequences of future climate change will be amplified by economic development and the tendency of populations to move into harm's way. Regardless of the relative weights of anthropogenic climate change and increased exposure (quantification is premature), rising uncertainty would complicate the fundamental actuarial and pricing processes that underlie well-functioning insurance markets.

The Globalization of Risk: One View of the Future

Most scenarios of climate change impacts are cast from the vantage point of the natural sciences with minimal examination of economic implications. Moreover, analysts often take an understandably simplified "stovepipe" approach by examining a specific type of event in isolation from the real-world mosaic of interrelated causes, vulnerabilities, and impacts.

The following business scenario—based on current socioeconomic trends and insurance market dynamics—explores an ensemble of events and impacts occurring simultaneously. The triggering events arise from the consequences of gradual anthropogenic climate change. Abrupt changes also anticipated in the longer term by many in the scientific community (*26*) are not included. The result is a plausible—and certainly neither a worst- nor best-case—rendition of what the future could bring.

In this scenario, weather-related property losses and business interruptions continue to rise at rates observed through the latter 20th century. The insured share increases, and underwriting becomes more problematic. Corporations face more environmentally related litigation (and associated insurance payouts), both as emitters of greenhouse gases and from noncompliance with new regulations (*27*).

A new class of losses emerges within the life and health branches of the insurance industry (*28*). These are driven by thermal extremes, reduced water quality and availability, elevated rates of vector-borne disease, air pollution, food poisoning, and injuries and mortalities from disasters and their associated mental health impacts (*11*, *29*, *30*). Other health consequences manifest in natural systems that directly or indirectly affect human systems, including coral-reef bleaching, agricultural diseases or other events that hamper food production, animal and livestock diseases, and pests such as pine beetle superinfestations. Mobilization of dust, smoke, and CO_2-linked aeroallergens (e.g., pollen and molds) exacerbates already high rates of asthma and other forms of respiratory disease.

The combined effect of increased losses, pressure on reserves, inflation of construction costs following natural disasters, and rising costs of risk capital result in a gradual increase in the number of years in which the industry is not profitable. A compounding effect arises from the continued destructive industry practice of underpricing risk and routinely allowing the core business to operate at a loss, relying instead on profits from their investments (also known as "cash-flow underwriting"). As occurred following the European windstorms of the 1990s (*7*), insurers periodically encounter liquidity problems when paying claims, forcing the sale of large blocks of securities, which, in turn, creates undesirable "knock-on" impacts in the broader financial markets. Outcomes are particularly bad in years when large catastrophe losses coincide with broader market downturns.

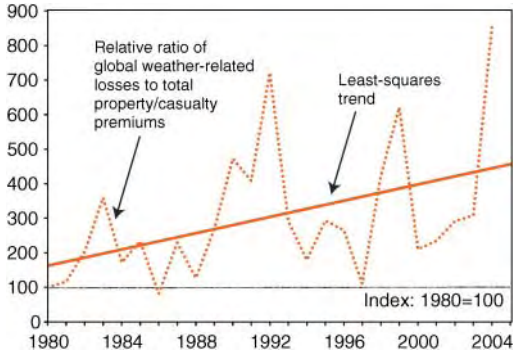


Fig. 3. Declining insurance industry capacity to absorb weather-related natural disasters. Curves show ratio of global weather-related property losses to total property/casualty premiums over the past quarter-century, indexed to average 1980 levels. [Sources as in Fig. 2]

Insurance operations in the developing world and economies in transition (the primary growth markets for insurance, already generating nearly $400 billion/year in premiums) are the most markedly affected. This arises from a combination of inferior disaster preparedness and recovery, more vulnerable infrastructure due to the lack or nonenforcement of building codes, high dependency on coastal and agricultural economic activities, and lack of funds to invest in disaster-resilient adaptation projects (*3*). Insurers also experience rising losses under political risk policies in these regions, as civil unrest and conflicts over food, water resources, and refugees manifest in the wake of natural disasters (*31*). Insurers from industrialized countries share these losses through their growing expansion into these emerging markets.

As insurability declines, insurers use traditional methods to reduce their exposures: increased premiums and deductibles, lowered limits, nonrenewals, and new exclusions. Although consumer demand for insurance increases at first, it evolves into reduced willingness to pay, and some shift from the use of insurance to alternatives such as weather derivatives.

As warned by the U.S. Government Accountability Office (*32*), private insurers encounter increasing difficulty in handling extreme weather events. As commercial insurability declines, demands emerge to expand existing government-provided insurance for flood and crop, and to assume new risks (e.g., wildfire and windstorm). Cash-strapped governments, however, find that claims interfere with balancing their budgets (*4*) and, in turn, limit their coverage (*33*), with the result that more ultimate losses are shifted back to the individuals and businesses affected by climate change. Compounding the problem, international aid for natural disasters continues its current decline as a percentage of donor-country GDP (*3*).

The impacts of climate change accelerate several forms of unrelated adverse structural change already under way in the insurance industry. This manifests as a rise in competition among insurers, consolidation due to the reduced viability of small firms, increased risk exposure by way of globalization, and a growing proportion of competing self-insurance and alternative risk transfer mechanisms. Better prepared for the impacts of climate change, European and Asian insurers capture market share from the United States.

Although the industry is not bankrupt, as some have suggested, an increasing number of firms do succumb to these losses, especially where solvency regulation is weak. In the United States, at least 7% of insurer bankruptcies are currently attributed to catastrophes (*34*). As the globe warms, climate change puts a chill on the insurance market. Insurance ceases to be the world's largest industry.

Policy Considerations

With human-induced climate change, the locus of Garrett Hardin's "tragedy of the commons" (*35*) as conceived in the 1960s has mushroomed from pastoral grazing lands to the global atmosphere. While his treatise was written before the issue was widely recognized, in a reprise three decades later Hardin reflected: "On the global scale, nations are abandoning not only the freedom of the seas, but the freedom of the atmosphere, which acts as a common sink for aerial garbage" (*36*). The relevance for insurers is twofold. They will inevitably experience some of the impacts and mirror back to society (through their selection and pricing of risk) some of the costs of externalities imposed on the climate commons. They may also become more proactive in formalizing social solidarity to prevent and, when necessary, endure and adapt to extreme events that individuals cannot manage independently, keeping the commons livable

and sustainable and the insurance business viable. This is the highest form of insurance, with roots in its centuries-old tradition of loss prevention. What happens in practice remains to be seen.

Insurance is a form of adaptive capacity for the impacts of climate change, although the sector itself must adapt in order to remain viable. It is incumbent on insurers, their regulators, and the policy community to develop a better grasp of the physical and business risks. A key issue is whether the meeting of commercial interests with the expectation that insurance serves a basic social function results in market failures under climate change. Related questions concern the nature and desirability of loss prevention, universal coverage, and profit-driven decisions to exclude coverage for certain individuals or hazards.

Can insurers extend their self-chosen historical role in addressing root causes (as founders of the first fire departments, building codes, and auto safety testing protocols) to one of preventing losses at a much larger scale, namely, the global climate? Should insurance regulators take a more active role in fulfilling their obligation to maintain insurance availability and affordability? How can the public be more effectively enlisted in loss prevention? What is government's role, and will it serve as "insurer of last resort"?

Although insurers first expressed concern about climate change more than three decades ago (*37*), fewer than one in a hundred appear to have seriously examined the business implications (*38*), and fewer still present their analyses in the open literature (*7*, *8*, *11*, *39*–*41*). This state of affairs heightens the likelihood of unanticipated adverse outcomes.

Disjointed modeling traditions and inconclusive attribution analyses hamper the industry's ability to assess weather-related risks and regulators' ability to safeguard both insurers and consumers (*1*). Insurers' weather-related loss models focus primarily on catastrophic events (to the exclusion of a broader array of small-scale events that have larger aggregate impacts), are predicated on extrapolating historical trends, and largely neglect life and health impacts. In contrast, the climate change community's models are future-focused, yield results not easily applied to business decision-making, and underestimate the physical and economic impacts of abrupt climate changes (*26*). These communities operate largely in isolation (*42*). An effort to bridge the gap, in the case of windstorms under climate change (*40*), yielded striking results. Predicted losses, technical prices (risk premiums), and capital requirements under a low-emissions scenario [525 parts per million by volume (ppmv) atmospheric CO_2 concentration by 2080] were one-fifth to one-eighth those under a high-emissions case (810 ppmv CO_2). The value of improved data and modeling is central, as evidenced by a shift in the industry (thanks in part to better models) toward accepting flood risks where they previously had been viewed as uninsurable (*43*).

Certain measures that integrate climate change mitigation and adaptation can simultaneously reduce insurance losses (*44*). Examples include protection of mangroves, reefs, and wetlands that buffer storm surge and wave risks. A host of energy-efficiency technologies also provide insurance loss-prevention benefits (*45*, *46*). A few insurers have sought to lead by example, e.g., Swiss Re has committed to becoming carbon-neutral in its operations and asset management, and a growing number have initiated enterprise-wide sustainability programs.

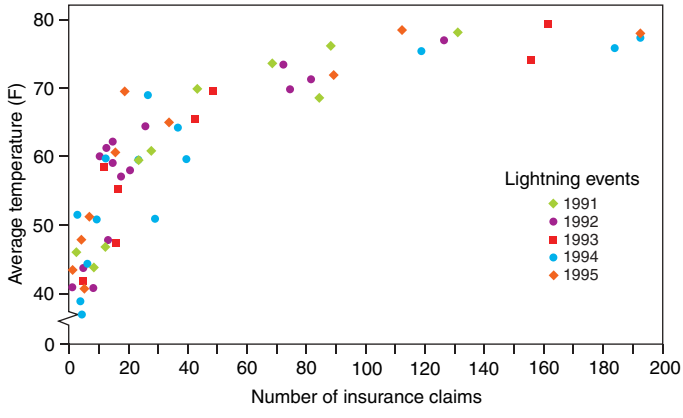


Fig. 4. Lightning-related insurance losses increase with temperature (*6*). Each symbol represents monthly losses in the continental United States.

Public-private partnerships are clearly essential (*11*, *43*, *47*). Insurers and governments have devised innovative means of spreading financial risk while fostering loss-prevention practices (*3*, *48*). Government's role can include research, education, emissions reductions, land-use planning, disaster preparedness, and provision of backstop reinsurance. Joint efforts to develop innovations such as microinsurance and microcredit can better serve the uninsured and fund disaster preparedness in the developing world. Such efforts should be coupled with a more prevention-oriented paradigm within the development and disaster relief communities.

Insurers will also find business opportunities in responding to climate change (*7*, *39*, *47*). These include new risk-management products for emissions reductions and loss-prevention technologies, analysis and advisory services, participation in adaptation activities, and increased demand for insurance itself.

Are insurers simply fomenting fear about climate change in order to sell more of their product or raise prices? This is doubtful. An industry thus motivated would be sounding a much louder alarm. Moreover, insurance regulators rarely allow price increases or expansion of reserves based on projected losses, and price competition is stiff. The prospects of forced reductions in coverage, loss of market share, periodic bankruptcies, eroded reputations, and regulator rejection of requests to withdraw from markets are material business risks that merit concern. Moreover, insurers hold major investments that may be vulnerable to climate change. A few insurers will no doubt be inappropriately opportunistic—and should be called to task for doing so—but those who have expressed concern are actively supporting climate change adaptation and mitigation, which will ultimately curb price increases.

The future role of insurance in helping society to cope with climate change is uncertain. Insurers may rise to the occasion and become more proactive players in improving the science and crafting responses. Or, they may retreat from oncoming risks, thereby shifting a greater burden to governments and individuals.

References and Notes

1. P. V. Vellinga *et al.*, in *Climate Change 2001: Impacts, Adaptation and Vulnerability*, J. J. McCarthy, O. F. Canziani, N. A. Leary, D. J. Dokken, K. S. White, Eds. (Cambridge Univ. Press, Cambridge, 2001), WG2, chap. 8.
2. These data are published annually in Swiss Re's *Sigma* series. For the purposes of this article, we use the terms "insurance" and "reinsurance" interchangeably.
3. E. Mills, "Insurance as an adaptation strategy for extreme weather events in developing countries and economies in transition: New opportunities for public-private partnerships" (Lawrence Berkeley National Laboratory Report No. 52220, Berkeley, CA, 2004).
4. D. Changnon, *Bull. Am. Meteorol. Soc.* 1231 (September 2003).
5. All-industry Research Advisory Council, "Catastrophic losses: How the insurance system would handle two $7 billion hurricanes" (American Institute for Chartered Property Casualty Underwriters, Insurance Institute of America, Insurance Research Council, Malvern, PA, 1986).
6. E. Mills, E. Lecomte, A. Peara, *J. Insur. Regul.* **21**, 1 (2002).
7. "A changing climate for insurance" (Association of British Insurers, London, 2004).
8. "Climate change futures: Executive roundtable" (Swiss Re, Centre for Global Dialogue, Ruchlekon, Switzerland, 2004).
9. C. Schar *et al.*, *Nature* **427**, 332 (2004).
10. In the wake of Hurricane Andrew, regulators limited or rejected insurers' requests for price increases or permission to cancel hundreds of thousands of policies. A decade later, in response to the four deductibles charged to some Florida homeowners following an intense year of hurricanes, regulators mandated reimbursement of 36,000 homeowners,

forbade insurers from canceling or nonrenewing victims, and required the instatement of "single-season" deductibles for the windstorm hazard.
11. *Weather Catastrophes and Climate Change* (Munchener Ruckversicherungs-Gesellschaft, Munich, 2005).
12. "Annual review of North American catastrophes 2004" (American Re, Princeton, NJ, 2005).
13. K. LaCommare, J. Eto, "Understanding the cost of power interruptions to U.S. electricity consumers" (Lawrence Berkeley National Laboratory Report No. 55718, Berkeley, CA, 2004).
14. Annual disturbance reports (1992–2002), North American Electric Reliability Council; available at www.nerc.com.
15. F. E. Nelson, O. A. Anisimov, N. I. Shiklomanov, *Nature* **410**, 889 (2001).
16. J. S. Fried, M. S. Torn, E. Mills, *Clim. Change* **64**, 169 (2004).
17. K. E. Kunkel, R. A. Pielke Jr., S. A. Changnon, *Bull. Am. Meteorol. Soc.* **80**, 1077 (1999).
18. S. A. Changnon, R. A. Pielke Jr., D. Changnon, R. T. Sylves, Pulwarty, *Bull. Am. Meteorol. Soc.* **81**, 437 (2000).
19. S. A. Changnon, M. Demissie, *Clim. Change* **32**, 411 (1996).
20. K. Zhang, B. C. Douglas, S. P. Leatherman, *Clim. Change* **64**, 41 (2005).
21. D. R. Easterling *et al.*, *Science* **289**, 2068 (2000).
22. T. R. Karl, K. E. Trenberth, *Science* **302**, 1719 (2003).
23. H. Kunreuther, E. Michel-Kerjan, "Insurance coping with global warming: Who will pay for large-scale risks associated with climate change?" Presentation to the Climate Decision Making Center, Carnegie Mellon University, 16 to 17 May 2005.
24. "Topics 2000—natural catastrophes: the current position" (Munich Re, Munich Reinsurance, Geoscience Research, Munich, 2000).
25. N. Reeve, R. Toumi, *Q. J. R. Meteorol. Soc.* **125**, 893 (1999).
26. R. B. Alley *et al.*, *Science* **10**, 1126 (2003).
27. M. R. Allen, R. Lord, *Nature* **432**, 551 (2004).
28. These issues are elaborated in a detailed scenario format in the forthcoming "Climate Change Futures" study, led by the Harvard Medical School's Center for Health and the Global Environment, sponsored by Swiss Re and the United Nations Development Programme.
29. "Climate change and human health: Risks and responses" (World Health Organization, Geneva, 2003).
30. P. Epstein, *Science* **285**, 347 (1999).
31. P. Schwartz, D. Randall, "An abrupt climate change scenario and its implications for United States national security" (Global Business Network, Emeryville, CA, 2003).
32. U.S. Government Accountability Office, *Catastrophe Risk: U.S. and European Approaches to Insure Natural Catastrophe and Terrorism Risks* (GAO-05-199, Washington, DC, 2005).
33. As a case in point, the risk of residential flooding in the United States is deemed largely uninsurable, which has given rise to a National Flood Insurance Program (NFIP), with more than 4.2 million policies in force, representing nearly $560 billion in coverage. The NFIP pays no more than $250,000 per loss per household and $500,000 for small businesses.
34. According to the AM Best's latest insolvency study. This may be a large underestimate because the values indicate "primary" cause of insolvency. An unspecified additional number of bankruptcies involve catastrophes as a contributing factor.
35. G. Hardin, *Science* **162**, 1243 (1968).
36. G. Hardin, *Science* **280**, 682 (1998).
37. "Flood inundation" (Munich Re, Munich Reinsurance Co., Munich, 1973).
38. These include insurance companies regularly convened over the past 10 years by the United Nation's Environment Programme. Current membership is 35 insurers from 17 countries. See www.unepfi.org/signatories/statements/ii.
39. "Opportunities and risks of climate change" (Swiss Re, Swiss Reinsurance Company, Zurich, Sigma 2/2002, 2002).
40. "Financial risks of climate change" (Association of British Insurers, London, 2005).
41. "Climate change and the financial sector: An agenda for action" (Allianz Group and World Wildlife Fund, Munich and Washington, DC, 2005).
42. F. W. Nutter, *J. Soc. Insur. Res.* 15 (Fall 1996).
43. I. Menzinger, C. Brauner, "Floods are insurable!" (Swiss Reinsurance Company, Zurich, 2002).
44. E. Mills, "Synergisms between climate change mitigation and adaptation: An insurance perspective," in *Mitigation and Adaptation Strategies for Global Change*, in press; available at http://eetd.lbl.gov/emills/PUBS/Insurance_A&M.html.
45. E. Mills, *J. Soc. Insur. Res.* 21 (Fall 1996).
46. E. Mills, *Energy Policy* **31**, 1257 (2003).
47. United Nations Environmental Programme and Innovest, *Climate Change and the Financial Services Industry* (UNEP, Geneva, 2002).
48. J. Linnerooth-Bayer, M. J. Mace, R. Verheyen, "Insurance-related actions and risk assessment in the context of the UNFCCC," Background paper for UNFCCC (United Nations Framework Convention on Climate Change) workshops (2003).
49. The research and analysis supporting this article was sponsored by the U.S. Department of Energy, the U.S. Environmental Protection Agency, and the U.S. Agency for International Development. Valuable discussions, review comments, and data were provided by S. Catavosky. A. Dlugolecki, P. Epstein, D. Grether, P. Höppe, I. Menzinger, R. Jones, E. Lecomte, J. McMahon, R. Roth, E. Saxon, J. Smith, M. Torn, and A. Wirtz.

10.1126/science.1112121

VIEWPOINT

Refocusing Disaster Aid

Joanne Linnerooth-Bayer, Reinhard Mechler, Georg Pflug

With new modeling techniques for estimating and pricing the risks of natural disasters, the donor community is now in a position to help the poor cope with the economic repercussions of disasters by assisting before they happen. Such assistance is possible with the advent of novel insurance instruments for transferring catastrophe risks to the global financial markets. Donor-supported risk-transfer programs not only would leverage limited disaster-aid budgets but also would free recipient countries from depending on the vagaries of postdisaster assistance. Both donors and recipients stand to gain, especially because the instruments can be closely coupled with preventive measures.

Postdisaster assistance for emergency relief and reconstruction, although important for humanitarian reasons, has failed to meet the needs of developing countries in reducing their exposure to disaster risks and ensuring sufficient funds to governments and individuals for financing the recovery process. Disasters continue to impose substantial human and economic losses on the developing world. In a sample of large natural disasters over the period 1980 to 2004, fatalities per event were higher by orders of magnitude in low- and middle-income countries compared with high-income countries; similarly, losses as a percentage of gross national income (GNI) were highly negatively correlated with per capita income (Fig. 1) (*1*). Despite evidence of large returns on investments in preventive measures (*2*), most assistance arrives after the disaster. Moreover, postdisaster aid discourages prevention because of the associated moral hazard: Governments and individuals, expecting support, have little incentive to invest in precautionary measures.

The donor community—financial institutions, international agencies, nongovernmental organizations, and donor governments—is recognizing the need to place more emphasis on programs that prevent disaster losses. There is less recognition, however, of the need to support risk-pooling and risk-transfer programs that assure readily available postdisaster funds for relief and reconstruction. Lacking sufficient funds, the follow-on costs of disasters can be extensive. For example, 4 years after the devastation of Hurricane Mitch in 1998, the gross domestic product (GDP) of Honduras was 6% below predisaster projections (*3*). Donor pledges of US$2.7 billion were considered exceptionally high but amounted to only about half of the estimated total reconstruction costs (*4*).

Governments, households, and businesses in poor countries cannot easily afford commercial insurance to cover their disaster risks. Whereas low-cost microinsurance for independent risks, such as funeral expenses, is now widely available in countries like Bangladesh and India, this is not the case for dependent risks that affect many communities at the same time. The cost of catastrophe insurance is usually substantially higher than the pure risk premium, mainly because of the insurer's cost of backup capital to cover dependent claims. Consequently, people can pay more for disaster insurance than their anticipated losses over the long term. For example, in the Caribbean region, catastrophe insurance

The authors are at the International Institute for Applied Systems Analysis, Laxenburg, Austria.

premiums were estimated to represent about 1.5% of GDP during the period 1970 to 1999, whereas average losses per annum (insured and uninsured) amounted to only about 0.5% of GDP (*5*). This helps explain why only 1% of households and businesses in low-income countries and only 3% in middle-income countries have catastrophe insurance coverage, compared with 30% in high-income countries (*1*). Instead of insurance, they rely on family and public support, which is not always forthcoming for catastrophes that affect whole regions or countries. Without support from insurance, family, or the government, disasters exacerbate poverty as victims take out high-interest loans (or default on existing loans), sell assets and livestock, or engage in low-risk, low-yield farming to lessen exposure to extreme events (*6*).

Developing countries cannot rely on donor aid to fill the financing gap. Humanitarian aid reported by the Organisation for Economic Co-operation and Development (OECD) Development Aid Committee is generally only a small percentage (usually under 10%) of disaster losses in recipient countries (*7*). [The $5.3 billion pledged after the 2004 Asian tsunami (*8*) was exceptional.] Moreover, promises can fall short of actual outlays. Two years after the 2001 earthquake in Gujarat, India, assistance from the central reserve fund and international sources had reached only 20% of original commitments (*9*). Because business as usual will likely not meet the postdisaster needs of developing countries, the donor community should consider refocusing disaster assistance to support risk-management programs that leverage aid with public and private contributions and that promote loss mitigation (*10*). Three examples demonstrate the potential for donor-supported risk-transfer programs.

The recently launched Turkish Catastrophe Insurance Pool (TCIP), which is similar to public-private insurance systems in France, the United States, Japan, and other developed countries, is the first of its kind to tackle the problem of insurance affordability in a middle-income developing country. Istanbul faces an estimated 0.41 probability of a severe earthquake over the 30-year period from 2004 to 2034 (*11*). In response to this risk, earthquake insurance policies are now obligatory for all property owners, who pay a risk-based fee to a privately administered public fund. To reduce the premiums and make the system viable, the World Bank provides backup capital for two layers of risk in the form of a loan with highly favorable conditions and contingent on the occurrence of a major disaster (*12*). Despite mandatory policies and 17% insurance penetration, the TCIP has not been fully successful in terms of public acceptance and implementation (and could have benefited from public participation in its design) (*13*). Still, it sets an important precedent as the first time an international financial institution has absorbed developing country risk. The donor community, too, could absorb risk in this way. Moving beyond the TCIP experience, it could provide support not just for risk transfer but for a comprehensive risk-management program, including preventive measures such as retrofitting buildings, educating the public, and enforcing land-use restrictions.

There are also innovative prospects for governments, which hold a large portfolio of infrastructure assets, to transfer their risks. Mexican authorities are planning to reinsure their national catastrophe relief and reconstruction fund (FONDEN) with a catastrophe bond. This is an instrument whereby the investor receives an above-market return when a specific catastrophe does not occur within a specified time (e.g., an earthquake of magnitude 7.0 or greater on the Richter scale in the vicinity of Mexico City over a 1-year period) but sacrifices interest or part of the principal after the event. The government's disaster risk is thus transferred to international financial markets that have many times the capacity of the reinsurance market (*14*). Although Mexico, a middle-income developing country, will finance the bond out of its own means, a similar, but donor-assisted, bond is an option for poorer countries.

Risk-transfer instruments are also emerging to address higher frequency, slower onset disasters such as droughts. In Ethiopia, where a drought resulted in the deaths of nearly one million people in 1984 (*15*), farmers may soon be able to purchase an index-based weather derivative. This is a contingent contract with a payoff determined by future weather events, such as a specified lack of precipitation measured at a weather station (*16*). Because of this physical trigger, there is no moral hazard; on the contrary, farmers have an incentive to reduce potential losses, for example, by diversifying their crops. A physical trigger also means that claims are not always fully correlated with losses, but this "basis risk" may be offset by the elimination of long and expensive claims settlements. Expedient payments will reduce the need for farmers to sell their assets and livestock to survive the aftermath of a disaster. However, despite low transaction costs, Ethiopian farmers have many other pressing needs and can ill afford to pay the costs of this hunger insurance. The affordability of the system will only be assured by the World Food Program, which has solicited donors to subsidize the premiums.

Such subsidized pilot schemes, if they can be scaled up to create a sufficiently diversified pool, hold considerable promise for the more than 40% of farmers in developing countries who face threats to their livelihoods from adverse weather (*17*). The donor community can help make these systems viable in two ways: subsidizing premiums as in Ethiopia and/or providing back-up capital to reduce the risks to private or public providers.

The implementation of these public-private risk-transfer programs is now feasible largely as a result of advances in computerized modeling that make it possible to better estimate and price low-probability extreme event risks for which there are limited historical data. Catastrophe models typically generate probabilistic losses by simulating stochastic events based on the geophysical characteristics of the hazard and combining the hazard data with analyses of exposure in terms of values at risk and vulnerability of assets. For example, the TCIP is based on a model that includes the tectonic or geophysical features of the country, seismic activity (modeled according to a compound Poisson process), and the intensity of ground-shaking. Apartment buildings have been mapped according to their vulnerability to shaking and the characteristics of the soil on which they are built (*18*). In addition, there has been important progress in the mathematics of extreme value theory and in the convergence of the theories of finance and insurance, rendering possible the pricing of exotic risk-transfer instruments such as weather derivatives and catastrophe bonds (*19*, *20*). Perhaps most important, countries are developing the capacity to build these models,

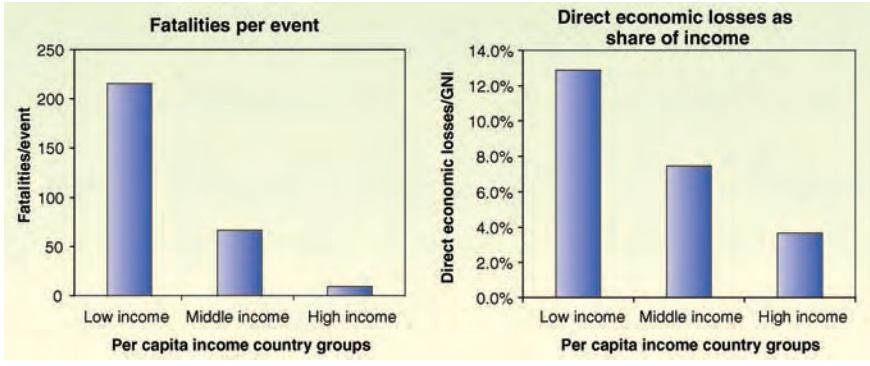


Fig. 1. Fatalities per event and direct economic losses as a share of national per capita income over the period 1980 to 2004. Country income groups according to World Bank classification using GNI per capita. Lower-middle and upper-middle income groups were combined.

which increases local knowledge about the risks and how to reduce them. On the basis of modeling and research by Turkish universities, the Turkish government is considering pilot programs to retrofit housing, schools, and other infrastructure.

If these pilot programs prove successful, they present compelling reasons for the donor community to follow the examples of the World Bank and the World Food Program and invest in risk-transfer instruments. By sharing responsibility with individuals and the state, donors leverage their limited budgets and substitute a calculable annual commitment to a risk-transfer system for the unpredictable granting of postdisaster aid. With donor-supported risk-transfer programs, developing country governments will rely less on debt financing and international donations, and assured funds for repairing critical infrastructure will attract foreign investment. Donor support will also provide poor households, farmers, and businesspersons with access to affordable means to spread risks spatially and temporally, which will secure their livelihoods and improve their creditworthiness. Most importantly, by making this assistance contingent on requirements or incentives for prevention as part of a comprehensive risk-management program, predisaster assistance can reduce the human and economic toll that disasters take on the poor. Even at extra cost, predisaster donor aid can thus be an efficient strategy if it reduces the long-term need for humanitarian disaster aid.

The need for assisting developing countries has become more pressing in light of the climate-change negotiations. As developed countries recognize that their greenhouse gas emissions can lead to increased intensity and frequency of weather extremes in the developing world (*21*), climate negotiators are seeking options for helping affected countries adapt. Specifically, the United Nations Framework Convention on Climate Change (UNFCCC) and the Kyoto Protocol call upon developed countries to consider actions, including insurance, to meet the specific needs and concerns of developing countries with respect to adverse climate impacts. Providing assistance to public-private risk-transfer programs such as those in Turkey, Mexico, and Ethiopia, is an innovative option to consider (*22*, *23*).

Recent initiatives to put risk-transfer programs into place, and the role of catastrophe models in making these programs possible, are remarkable. Still, there are many challenges for implementing these programs on a large scale and ensuring that they genuinely provide affordable security to the poor. To begin, the science underlying the models and risk estimates must be independent, transparent, and viewed as reliable by insurers, investors, and donors. Despite advancements in data collection and verification with satellite technology, changes in climate, urbanization, and land-use practices create large uncertainties in estimating longer term risks (catastrophe insurance arrangements are usually short term) and add to the reluctance of the private sector to invest in risk-transfer instruments (*24*).

Beyond the formidable scientific issues, a challenge for the international donor community is to promote good governance and sound regulatory practices as prerequisites for any risk-transfer program. As cases in point, premium payments for the TCIP are placed in a segregated account, and investors in catastrophe bonds are assured payment from an internationally controlled special-purpose vehicle. Both are legally inaccessible to the respective governments. Still, corrupt practices are a risk even to clear contractual arrangements. This political risk, however, should be compared to that of ad hoc postdisaster assistance, which is more difficult to monitor and often diverted from its intended purpose.

Another challenge is overcoming the myopia of individuals and governments and their reluctance to invest in protection against low-probability disasters. Subsidizing developing country insurance programs, it can be argued, will distort prices and create the wrong signals for avoiding risk exposure. This argument, however, is hardly relevant for poor communities, which have few affordable options for relocating or otherwise reducing their exposure to disaster risk, and, lacking subsidized insurance instruments, will continue to rely on international aid. The argument against subsidies, however, underlines the importance of explicitly tying predisaster support to affordable loss prevention and phasing out subsidies as countries develop.

As these early risk-transfer schemes illustrate, the purpose of refocusing disaster aid is not to replace it with unaffordable private insurance but rather to complement postdisaster humanitarian aid with predisaster support of risk-management programs that combine prevention and risk transfer. A strong commitment on the part of the donor community and climate policy negotiators in providing expertise, capacity building, and, above all, financial support will be necessary if these programs are to make a wide-scale difference to the poor.

References and Notes

1. Munich Re NatCatSERVICE, *Natural Disasters According to Country Income Groups 1980-2004* (Munich Re, Munich, 2005).
2. R. Mechler, "Cost-benefit analysis of natural disaster risk management in developing countries," working paper (Deutsche Gesellschaft für Technische Zusammenarbeit, in press).
3. R. Mechler, *Natural Disaster Risk Management and Financing Disaster Losses in Developing Countries* (Verlag für Versicherungswissenschaft, Karlsruhe, 2004).
4. J. Telford *et al.*, "Learning lessons from disaster recovery: The case of Honduras," Disaster Risk Management Working Paper Series No. 8 (World Bank, Washington, DC, 2004).
5. P. Auffret, "Catastrophe insurance market in the Caribbean region," World Bank Policy Research Working Paper No. 2963 (World Bank Group, Washington, DC, 2003).
6. P. Varangis, P. Skees, J. R. Barnett, in *Climate Risk and the Weather Market*, B. Dischel, Ed. (Risk Books, London, 2005), pp. 279–294.
7. J. Linnerooth-Bayer, A. Amendola, *Geneva Pap. Risk Ins.* **25**, 203 (2000).
8. Reuters Foundation AlertNet, www.alertnet.org/thefacts/aidtracker/.
9. World Bank, *Financing Rapid Onset Natural Disaster Losses in India: A Risk Management Approach* (World Bank Group, Washington, DC, 2003).
10. UNISDR (United Nations International Strategy for Disaster Reduction), Hyogo Declaration. World Conference on Disaster Reduction (Kobe, Japan, 2005).
11. T. Parsons, *J. Geophys. Res.* **109**, B05304, doi:10.1029/2003JB002667 (2004).
12. E. Gurenko, in *Catastrophe Risk and Reinsurance: A Country Risk Management Perspective*, E. Gurenko, Ed. (Risk Books, London, 2004), pp. 3–16.
13. In Hungary, a participatory process provided important insights to the design of a national insurance system (*25*).
14. J. Pollner, *Managing Catastrophic Risks Using Alternative Risk Financing and Insurance Pooling Mechanisms* (World Bank, Washington DC, 2000).
15. J. Morris, "Can insurance break Ethiopia's vicious cycle of hunger?" (FT.com Financial Times, 10 May 2005).
16. U. Hess, K. Richter, A. Stoppa, in *Climate Risk and the Weather Market*, B. Dischel, Ed. (Risk Books, London, 2005), pp. 295–310.
17. World Bank, "Commodity and Weather Risk Management Programs to be Expanded" (World Bank News & Broadcast, 24 May 2005). http://web.worldbank.org/WBSITE/EXTERNAL/NEWS/0,,contentMDK:20513603~menuPK:34459~pagePK:34370~piPK:34424~theSitePK:4607,00.html.
18. D. Kuzak, in *Catastrophe Risk and Reinsurance: A Country Risk Management Perspective*, E. Gurenko, Ed. (Risk Books, London, 2004), pp. 41–64.
19. P. Embrechts, C. Klüppelberg, T. Mikosch, *Modeling Extreme Events* (Springer Verlag, Berlin, 1997).
20. H. Geman, Ed., *Insurance and Weather Derivatives: From Exotic Options to Exotic Underlyings* (Risk Books, London, 1999).
21. IPCC, *Climate Change 2001: The Scientific Basis. Contribution of Working Group 1 to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, 2001).
22. J. Linnerooth-Bayer, M. J. Mace, R. Verheyen, "Insurance-related actions and risk assessment in the context of the UNFCCC," background paper prepared for the UNFCCC Secretariat (UNFCCC, Bonn, 2003). http://unfccc.int/meetings/workshops/other_meetings/items/1043.php.
23. J. Linnerooth-Bayer, R. Mechler, *Financing Disaster Risks in Developing and Emerging-Economy Countries*, OECD Conference on Catastrophic Risks and Insurance, Paris, 22 to 23 November 2004.
24. E. Mills, *Science* **309**, 1040 (2005).
25. J. Linnerooth-Bayer, A. Vari, *Clumsy Solutions for a Complex World*, in preparation.
26. We are grateful for helpful comments provided by E. Bergschneider, L. MacKellar, W. Sanderson, and M. Thompson, as well as four anonymous referees.

10.1126/science.1116783

Soil Invertebrates Disrupt Carbon Flow Through Fungal Networks

David Johnson,[1]*† Martin Krsek,[2] Elizabeth M. H. Wellington,[2] Andrew W. Stott,[3] Lisa Cole,[4] Richard D. Bardgett,[5] David J. Read,[6] Jonathan R. Leake[6]*

A major proportion of the flux of carbon (C) from plants to soil moves rapidly through networks of symbiotic, arbuscular mycorrhizal (AM) fungi (*1*), which colonize around 80% of land plants. Mycorrhizal mycelia extends from plant roots to produce a distinct "mycorrhizosphere" and a nutrient-absorbing network, which, in grassland, can reach up to 100 m in length per gram of soil (*2*). Increasing awareness of the importance of AM mycelial networks for below-ground C flow has highlighted the need to investigate the extent to which these fragile pathways are susceptible to disruption by abundant fungal-feeding invertebrates (*3*). Here we address the affect of natural densities of Collembola in their native grassland habitat on C flow through intact mycelial networks. We combined field-based manipulations of AM external mycelium and fungal-feeding microarthropod populations with in situ $^{13}CO_2$ pulse labeling (*4*) in upland, seminatural grassland in Scotland, UK.

Mesh cores, into which mycorrhizal mycelium could grow from the surrounding plant roots and in which collembolans could be "caged," were filled with defaunated soil and inserted into replicate plots (*4*). After 6 weeks for mycelial colonization, 20 individuals of the collembolan *Protaphorura armata* were added to half the cores; the remaining cores served as controls (Fig. 1). After a further 4 weeks, the swards of the plots containing the cores were pulse labeled with $^{13}CO_2$ for 7 hours (*4*).

The numbers of *P. armata* surviving in the mesh cores to which they were added ranged from 2 to 8 (SEM, 5 ± 0.58). The relatively low survivorship of the collembolans, perhaps in part a result of the extraction procedure (*4*), resulted in a density of 14,100 m^{-2}, which is within the typical range for temperate grassland (5,000 to 50,000 m^{-2}). The presence of *P. armata* reduced by 32% the atom excess $\%^{13}C$ (the relative proportion of ^{13}C above that found naturally) of respiration from the mycorrhizosphere (Fig. 1) ($P = 0.014$).

The addition of the collembolans had no effect on the natural abundance $\delta^{13}C\text{-}CO_2$ (the relative proportion of ^{13}C found naturally) from mesh cores (Fig. 1) ($P = 0.55$), so their effects on ^{13}C enrichment of respiration after pulse labeling must reflect disruption to the pathway of C flux into the mycorrhizosphere. This disruption could be a result of either consumption of mycelium or incidental damage. We found that the phospholipid fatty acid (PLFA) 16:1ω5 was significantly (Student's $t = 7.09$, $P = 0.006$) enriched in ^{13}C in the collembola-free mycorrhizosphere after pulse labeling, whereas in the presence of collembola, this PLFA was not significantly enriched (Student's $t = 2.43$, $P = 0.093$). Because 16:1ω5 has high specificity to AM fungi (*5*), this result suggests that the reduction in atom excess $\%^{13}C$ of respiration as a result of the activities of the collembola is primarily attributable to disruption of mycorrhizal rather than saprotrophic fungal mycelium. This supports observations that have identified fragments of AM mycelium in the guts of some Collembola (*6*).

Given the importance of mycorrhizal mycelial networks for the establishment, diversity, nutrition, and productivity of plant communities (*7*), their disruption by soil invertebrates may have a wide range of additional effects that need to be investigated. Our results highlight the need to redouble efforts to quantify mycorrhizosphere C fluxes in the field and its contribution to soil food webs. There is now a clear case for inclusion of C flow through the mycorrhizosphere in empirical and theoretical studies of C cycling in natural environments and for further work to estimate the effect of soil mesofauna on the total C flux from soil.

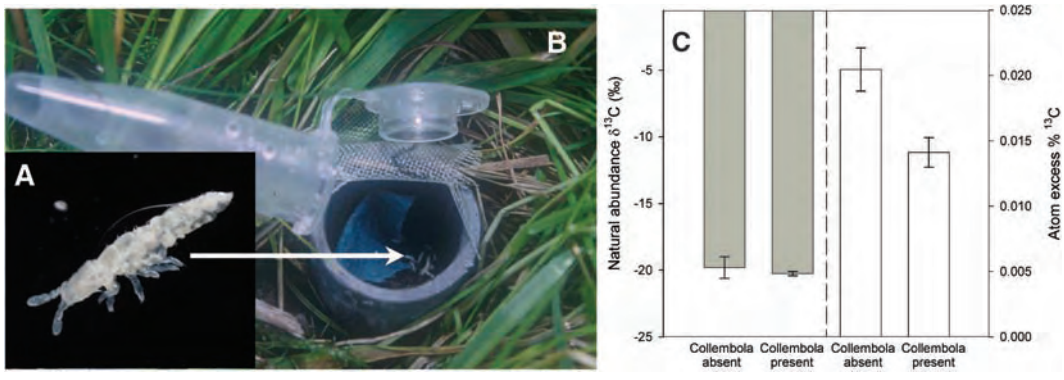


Fig. 1. (**A**) Twenty collembolans (*P. armata*) were freshly extracted and added to (**B**) mesh-walled root exclusion cores containing nonsterile defaunated soil that had been inserted into field plots in upland grassland (*4*). (**C**) Effect of the presence and absence of natural densities of *P. armata* on the natural abundance isotopic signature (shaded bars, not significantly different; $P = 0.55$) and on the atom excess $\%^{13}C$ released from mesh cores containing root-free mycorrhizosphere soil 16 hours after pulse labeling (open bars, significantly different; $P = 0.014$). The CO_2 released from the mesh cores was sampled within 90 min of their removal from the field plots (*4*) and the isotopic signatures analyzed by paired *t* test (± SEM).

References and Notes

1. P. L. Staddon, C. B. Ramsay, N. Ostle, P. Ineson, A. H. Fitter, *Science* **300**, 1138 (2003).
2. J. R. Leake *et al.*, *Can. J. Bot.* **82**, 1016 (2004).
3. A. Gange, *Trends Ecol. Evol.* **15**, 369 (2000).
4. Materials and methods are available as supporting material on *Science* Online.
5. P. A. Olsson, E. Bååth, I. Jakobsen, B. Söderström, *Mycol. Res.* **99**, 623 (1995).
6. T. P. McGonigle, A. H. Fitter, *Proc. R. Soc. Edinb. Sect. B* **94**, 2 (1988).
7. M. G. A. Van der Heijden *et al.*, *Nature* **396**, 69 (1998).
8. We thank N. Ostle and N. P. McNamara for supplying the pulse label and I. Johnson, C. Macdonald, and G. Burt-Smith for help with fieldwork. Supported by the NERC's Soil Biodiversity initiative. D.J. is currently supported by a NERC Advanced Fellowship.

Supporting Online Material
www.sciencemag.org/cgi/content/full/309/5737/1047/DC1
Materials and Methods
References and Notes

12 May 2005; accepted 24 June 2005
10.1126/science.1114769

[1]School of Biological Sciences, University of Aberdeen, Aberdeen AB24 3UU, UK. [2]Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK. [3]Natural Environment Research Council (NERC) Life Sciences Mass Spectrometry Facility, Lancaster Environment Centre, Lancaster LA1 4AP, UK. [4]Centre for Ecology and Hydrology, Banchory AB31 4BW, UK. [5]Institute of Environmental and Natural Sciences, University of Lancaster, Lancaster LA1 4YQ, UK. [6]Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: D.Johnson@abdn.ac.uk

Atomic-Scale Sources and Mechanism of Nanoscale Electronic Disorder in $Bi_2Sr_2CaCu_2O_{8+\delta}$

K. McElroy,[1,2] Jinho Lee,[1] J. A. Slezak,[1] D.-H. Lee,[2] H. Eisaki,[3] S. Uchida,[4] J. C. Davis[1]*

The randomness of dopant atom distributions in cuprate high–critical temperature superconductors has long been suspected to cause nanoscale electronic disorder. In the superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$, we identified populations of atomic-scale impurity states whose spatial densities follow closely those of the oxygen dopant atoms. We found that the impurity-state locations are strongly correlated with all manifestations of the nanoscale electronic disorder. This disorder occurs via an unanticipated mechanism exhibiting high-energy spectral weight shifts, with associated strong superconducting coherence peak suppression but very weak scattering of low-energy quasi-particles.

Important new electronic materials have been created by chemically doping their inert parent compounds with charge carriers. Examples include organic conductors, cuprate high-temperature superconductors, manganites that display colossal magnetoresistance, cobaltates that display large thermoelectric powers, and magnetic semiconductors. In such chemical doping schemes, a critical challenge is to generate the desired electronic properties without also creating destructive electronic disorder due to the random distribution of dopant atoms.

Cuprate high-T_c superconductivity (HTSC) is one of the most remarkable examples of this situation. In hole-doped cuprates (those with the highest critical temperature T_c), superconductivity appears when electron acceptor atoms are introduced into the parent insulators (*1*). This process almost always creates random distributions of differently charged dopant atoms near the critical CuO_2 crystal planes. These are potential sources of electronic and crystalline disorder (*2*) and T_c suppression (*3*, *4*). Indeed, dopant atoms may have two diametrically opposite influences on HTSC, being destructive at the atomic scale but supportive globally (via enhanced carrier densities). Thus, a key challenge in cuprate studies has been to determine directly how dopant atoms influence the superconducting electronic structure at the atomic scale.

Spectroscopic imaging scanning tunneling microscopy (SI-STM) studies reveal nanoscale electronic disorder in many cuprates. For example, $YBa_2Cu_3O_{6-\delta}$ exhibits local density of states (LDOS) modulations at the CuO plane (*5*). In $Na_xCa_{2-x}CuO_2Cl_2$, topographic imaging shows both random nanoscale electronic disorder (*6*) and disorder related to the 4×4 electronic "checkerboard" of nondispersive LDOS modulations (*7*). $Bi_2Sr_2CaCu_2O_{8+\delta}$ (Bi-2212) exhibits several signatures of electronic disorder, including subgap dispersive LDOS modulations (*8*), disordered regions (diameter ~3 nm) in energy-gap images $\Delta(\vec{r})$ (*9*–*11*), and nanoscale disorder in low-bias topographic images $z(\vec{r})$ (*11*). Although dopant disorder was suspected to cause these effects, this was impossible to verify directly because simultaneous atomic-resolution images of the dopant atom locations and the superconducting electronic structure were unattainable.

Locating the atomic-scale sources of electronic disorder. Here we focus on Bi-2212, a cuprate superconductor that is hole-doped by adding a number δ of nonstoichiometric oxygen atoms per unit cell. The floating zone–grown crystals, prepared with variable oxygen dopant densities, can be cleaved to reveal the BiO layer (Fig. 1A). Although analyses of x-ray and neutron scattering studies imply that the oxygen dopant atoms may be located either within or between BiO layers (*12*, *13*), their exact locations have never been determined. To attempt location of oxygen dopants by SI-STM, one needs to identify an electronic impurity state characteristic of dopant atoms only. The stoichiometric oxygen atoms are invisible to low-bias SI-STM because their contributions to the BiO band structures occur far from the Fermi energy. By contrast, each nonstoichiometric oxygen dopant atom is a $-2e$ charged impurity, so they might be expected to generate an electronic impurity state somewhat below the Fermi level. To search for such dopant-induced impurity states, we acquired LDOS($\vec{r}$,$E = eV$) images with large field of view over a range of energies -1.2 eV $< E <$ -0.15 eV. At typical tunnel junction resistance $R_J = 10^9$ ohms with $V \sim 1$ V, the electric field exceeds $\sim 10^9$ V/m and can rupture the BiO surface; therefore, new techniques for LDOS imaging with R_J up to 10^{11} ohms were developed for these studies.

Figure 1A shows a typical $z(\vec{r})$ topograph with sample bias $V_B = -0.9$ V. Figure 1B, which shows the simultaneous LDOS($\vec{r}$,E) image at energy $E = -0.9 \pm 0.05$ eV, reveals a random distribution of bright maxima in the LDOS, each with radius ~4 Å. The inset of Fig. 1A compares the shape of the high-voltage spectrum at one of the bright features in Fig. 1B with that of a spectrum typical of regions far from them; we see that the contrast in Fig. 1B occurs because each bright region exhibits a broad peak (arrow in Fig. 1A) in the spectrum at $V = -0.96$ V.

By cross-correlating the simultaneous topographic and LDOS($V = -0.96$ V) images (e.g., Fig. 1A with Fig. 1B), we determined that the "–0.96 eV features" occur at a distance of ~1.9 Å away from the nearest Bi along the BiO bond direction but in an unknown plane (inset, Fig. 1B). The precision of this determination is not high because of crystal spatial disorder and the extended nature (diameter ~8 Å) of the impurity state. One possible explanation for these features is a dopant atom residing above the BiO layer. This scenario, although not ruled out, is not supported by observations of adatoms at these locations. Another possibility is a vacancy at the stoichiometric oxygen site in the BiO layer, but here topography does not reveal the anticipated displacements of Bi atoms. A well-known electronic defect (*4*) is substitution of a Bi atom on the Sr site, but this cannot explain the –0.96 eV impurity state because it is positively charged (Bi^{3+} on the Sr^{2+} site), so its impurity state should exist above the Fermi level; in fact it is observed at $E = +1.7$ eV (*9*). A final possibility is that, at whichever site within the crystal the O^{-2} dopant atom resides, it generates a local electronic impurity state somewhere below the Fermi level.

To test whether the "–0.96 V features" could be such impurity states, we repeated the LDOS($V = -0.96$ V) imaging experiments on several samples with different oxygen dopant densities. Typical results are shown in Fig. 2, where the locations of spectra exhibiting –0.96 V peaks are marked as white dots superimposed on the simultaneously acquired gap maps $\Delta(\vec{r})$. The hole density p per CuO_2 decreases from Fig. 2A to Fig. 2C, as evidenced by the increasing average gap magnitude $\overline{\Delta}$; in break junction studies of Bi-2212, $\overline{\Delta} \approx [75 - 275\ (\pm 10\%) \times p]$ meV (*14*). From images of the –0.96 eV features, we

[1]Laboratory of Atomic and Solid State Physics, Department of Physics, Cornell University, Ithaca, NY 14850, USA. [2]Department of Physics, University of California, Berkeley, CA 94720, USA. [3]National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Central 2, Umezono, Tsukuba, Ibaraki 305-8568, Japan. [4]Department of Physics, University of Tokyo, Tokyo 113-8656, Japan.

*To whom correspondence should be addressed. E-mail: jcdavis@ccmr.cornell.edu

find that their density per CuO_2 plaquette, n, also diminishes from Fig. 2A to Fig. 2C and is related to the average gap magnitude $\bar{\Delta} \approx [75 - 600\ (\pm 15\%) \times n]$ meV. These data are all consistent with every –0.96 eV state being associated with the contribution of ~2 holes to the CuO_2 plane. It is then reasonable to postulate that each is due to a single oxygen dopant atom, and hereafter we refer to them provisionally as the "dopant defects." These –0.96 eV impurity states, independent of their microscopic cause or electronic structure, occur at the primary sources of the nanoscale electronic disorder (see below).

Superconducting electronic disorder phenomena in Bi-2212. The simplest type of electronic disorder in Bi-2212 appears in LDOS($\vec{r}$,E) images generated from tip-sample differential tunneling conductance $g(\vec{r},V)$ measurements:

$$\mathrm{LDOS}(\vec{r}, E = eV) \propto g(\vec{r}, V) \equiv dI/dV\big|_{\vec{r},V} \quad (1)$$

Differential conductance images revealed weak quasi-periodic modulations exhibiting up to 16 inequivalent sets of energy-dispersive wave vectors in $g(\vec{q},E)$, the Fourier transform magnitude of $g(\vec{r},E)$ (*8*). When analyzed as scattering-induced quasi-particle interference patterns (*15*–*17*), these sets of dispersing $\vec{q}$-vectors yielded elements of the Fermi surface and momentum space $\Delta(\vec{k})$ in excellent agreement with photoemission (*8*, *18*, *19*). However, the identity of the scattering centers that generate these quasi-particle interference patterns was unknown.

Intense disorder is also universally seen in atomically resolved images of the superconducting energy gap $\Delta(\vec{r})$ (*9*–*11*, *20*–*23*). Gap values range from below 20 meV to above 65 meV within adjacent nanoregions; the data shown in Fig. 2, A to C, are typical. Explaining these $\Delta(\vec{r})$ phenomena has become a key challenge in cuprate studies.

The disorder seen in low-bias (–150 mV < V_B < 0) topographic images of Bi-2212 is also exceptionally strong (*11*). For example, the low-bias topograph in Fig. 3B (with the ~2.6 nm wavelength surface undulations removed by Fourier filtering) reveals the disorder contrast as dark and light nanoregions. STM topographic imaging actually measures the vertical tip displacement $z(\vec{r},V_B)$ required to maintain constant tunnel current $I(V_B)$ for sample bias voltage V_B. With z_0 as the exponential tunneling factor, the current at V_B is given by

$$I(V_B) \propto \exp[-z(\vec{r}, V_B)/z_0] \times \int_0^{eV_B} \mathrm{LDOS}(E, \vec{r})\,dE \quad (2)$$

$$\Rightarrow z(\vec{r}, V_B) + C = z_0 \ln \int_0^{eV_B} \mathrm{LDOS}(E, \vec{r})\,dE \quad (3)$$

In this simple single-particle picture, the spatial charge density variations within states up to $E = eV_B$ would then be

$$\delta\rho(\vec{r})\Big|_0^{eV_B} = \delta\left[e\int_0^{eV_B} \mathrm{LDOS}(E, \vec{r})\,dE\right] \quad (4)$$

For this reason, the disorder seen in low-bias topographs $z(\vec{r}, -150\ \mathrm{mV} < V_B < 0)$ has been interpreted via Eqs. 3 and 4 as a logarithmic image of charge density disorder (*11*, *24*).

Modeling the cause of nanoscale electronic disorder in cuprates. Nanoscale electronic disorder effects in cuprates (*5*–*11*, *20*–*23*) pose critically important questions. Are there really intense nanoscale CuO_2 plane charge density variations in near-optimal high-T_c superconductors? Is the observed electronic disorder due to some form of electronic self-organization or to dopant disorder? If the latter, what is the microscopic mechanism? Is it scattering from dopant atoms that produces the quasi-particle interference modulations (*8*) and superconducting coherence peak suppression (*10*, *22*, *23*)? Which specific dopant atom or defect locations most strongly diminish T_c (*3*, *4*)?

Theoretical studies of the effects of dopant atoms explore several hypotheses. In one such model (*24*–*26*), poor electrostatic screening of the –2e charge on each dopant atom produces a repulsive potential for electrons, causing holes to accumulate at that location. This local increase in hole density produces reduced energy-gap values near the dopant atoms (*24*–*26*). Thus, in such a "hole accumulation" model, the disorder in $\Delta(\vec{r})$ is created electrostatically by the dopant atoms via hole density variations in the CuO_2 plane. The second class of model (*27*) holds that, because of their displacement from the CuO_2 plane, the dopant oxygens produce only small-angle scattering, which is nonetheless sufficient to result in both the observed quasi-particle interference modulations and the gap disorder (*27*). A third model is that the dopants generate local crystalline stress/strain at the nanoscale (*2*, *3*, *28*, *29*). In

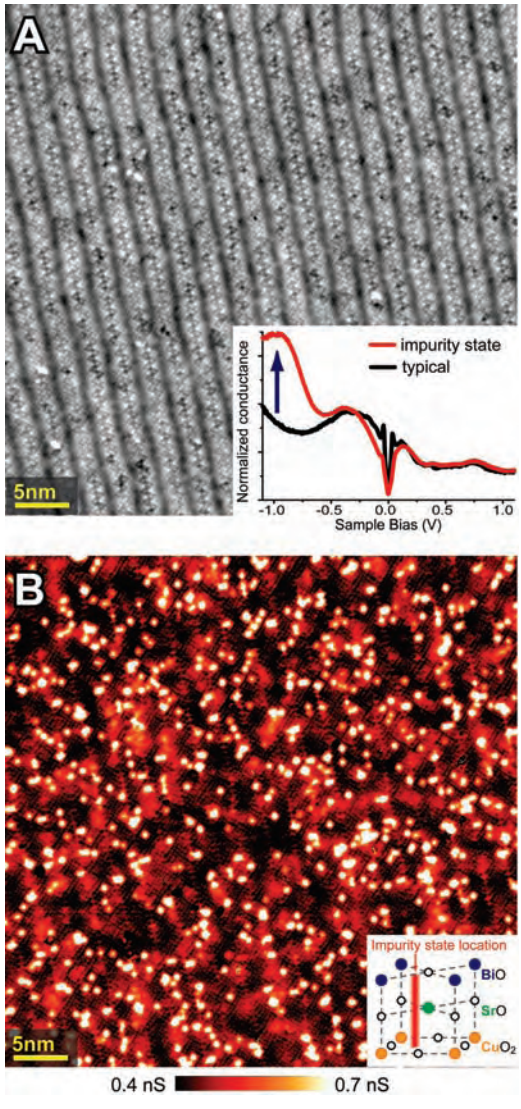


Fig. 1. (**A**) Atomic-resolution topographic image of the BiO surface of Bi-2212 taken at V_B = –0.9 V and $R_J = 3 \times 10^{10}$ ohms. The inset compares the shape of a high-voltage spectrum at one of the bright features in (B) with that of a spectrum typical of regions far from any bright feature (spectra were taken at different junction resistance and have different normalizations). (**B**) An LDOS(E = –0.96 eV,$\vec{r}$) ≡ O($\vec{r}$) image [simultaneously acquired with (A)] of the differential tunneling conductance measured at V = –0.9 V with a modulation amplitude of 50 mV. It shows about 600 atomic-scale regions of high differential conductance, which occur because there is a peak in the differential conductance spectrum near V = –0.96 V at each of these locations. This is the only energy in our window where such effects were detected. Examples of the high-voltage spectra on and off these bright regions are shown in the inset of (A); an arrow indicates the peak in tunnel conductance. Shown schematically in the inset of (B) is the line parallel to the c axis along which we believe the dopant defect is located. It is not possible to determine which layer of the crystal it lies in.

this picture, the local electronic disorder occurs because various hopping matrix elements in the crystalline unit cell are altered by crystal distortions near dopants. It seemed reasonable that the exact cause of cuprate nanoscale electronic disorder could contain some contributions from all these, and perhaps other, effects.

Correlating dopant defect locations with electronic disorder. To explore these issues, we acquired simultaneous atomically resolved (i) topographs $z(\vec{r},V_B)$, (ii) LDOS$(E,\vec{r})$ images, (iii) gap maps $\Delta(\vec{r})$, and (iv) LDOS(E = –0.96 eV,$\vec{r}$) ≡ O($\vec{r}$) dopant-defect images for each sample (more than 3 × 10^5 individual spectra for this study). To analyze such data, we define the normalized cross-correlation $C_{f,g}(\vec{R})$ between two-dimensional images $f(\vec{r})$ and $g(\vec{r})$ as

$$C_{f,g}(\vec{R}) = \frac{\int \left[f(\vec{r}) - \bar{f}\right] \times \left[g(\vec{r}+\vec{R}) - \bar{g}\right] d^2r}{\sqrt{A_{f,f}(0)A_{g,g}(0)}} \quad (5)$$

where

$$A_{f,f}(\vec{R}) = \int \left[f(\vec{r}) - \bar{f}\right] \times \left[f(\vec{r}+\vec{R}) - \bar{f}\right] d^2r \quad (6)$$

For simplicity, we designate $C_{f,g}(\vec{R} = 0)$ as $C[f{:}g]$. Identical images $f(\vec{r}) = g(\vec{r})$ yield $C[f{:}g] = 1$, an image and its negative $f(\vec{r}) = -g(\vec{r})$ yield $C[f{:}g] = -1$, and $C[f{:}g] = 0$ for no local correlation. We calculate a series of cross-correlations between the simultaneously measured $z(\vec{r},V_B)$, LDOS$(E,\vec{r})$, $\Delta(\vec{r})$, and O($\vec{r}$) images, yielding similar results at all dopings studied. A typical set of $C_{f,g}(\vec{R})$ data averaged azimuthally (Fig. 3A) immediately reveals the strong spatial correlations between dopant defects and the electronic disorder.

Important deductions emerge from more detailed analysis of these cross-correlations. It is well known that low-bias topographic disorder and gap maps are anticorrelated (*11*), and Fig. 3A shows that our measured cross-correlation coefficient between $\Delta(\vec{r})$ and $z(\vec{r},V_B = -100$ mV), $C[\Delta(\vec{r}){:}z(\vec{r},V_B = -100$ mV)] ≈ –0.4, in good agreement. But we can now demonstrate that this effect is indeed generated by dopant defects, because (Fig. 2, A to C; Fig. 3A) strong correlations exist between dopant defect images O($\vec{r}$) and both (i) low-bias topographic disorder $z(\vec{r},V_B = -100$ mV), $C[\mathrm{O}(\vec{r}){:}z(\vec{r},V_B = -100$ mV)] ≈ –0.25, and (ii) simultaneous gap maps $\Delta(\vec{r})$, $C[\mathrm{O}(\vec{r}){:}\Delta(\vec{r})] \approx +0.3$. But these correlations are inconsistent with dopant-induced hole accumulation models (*24*–*26*) because the sign of the correlation $C[\mathrm{O}(\vec{r}){:}\Delta(\vec{r})]$ is positive. This means that instead of diminished gap magnitudes, we find predominantly increased gap magnitudes near each dopant defect (Fig. 2, A to C; Fig. 3A). We find that the sign of $C[\mathrm{O}(\vec{r}){:}z(\vec{r},V_B)]$ is negative for both signs of sample bias. This means that low-bias $z(\vec{r},V_B)$ is almost equally diminished for states above and below the Fermi energy. Even more important, as V_B is increased, correlations with pairs of topographs of opposite V_B remain identical, but the topographic $z(\vec{r},V_B)$ disorder itself begins to moderate near $V_B = \pm 0.25$ V and is greatly diminished above $V_B = \pm 0.60$ V (Fig. 3C). Consistent with this, Fig. 3A shows that $C[\mathrm{O}(\vec{r}){:}z(\vec{r},V_B = -900$ mV)] ≈ 0 because disorder in $z(\vec{r},V_B = \pm 0.9$ V) is very weak. All these data indicate that, when integrated over states in the range $V_B \approx \pm 0.9$ V, variations in $\int_0^{eV_B}$ LDOS$(E,\vec{r})dE$ are significantly weaker than had been originally suspected [as had been anticipated from nuclear magnetic resonance studies (*30*, *31*)]. Within the single-particle picture, we estimate that root mean square charge density fluctuations are $\delta\rho(\vec{r})/\rho \leq 10\%$ in the states up to $V_B \approx \pm 0.9$ V.

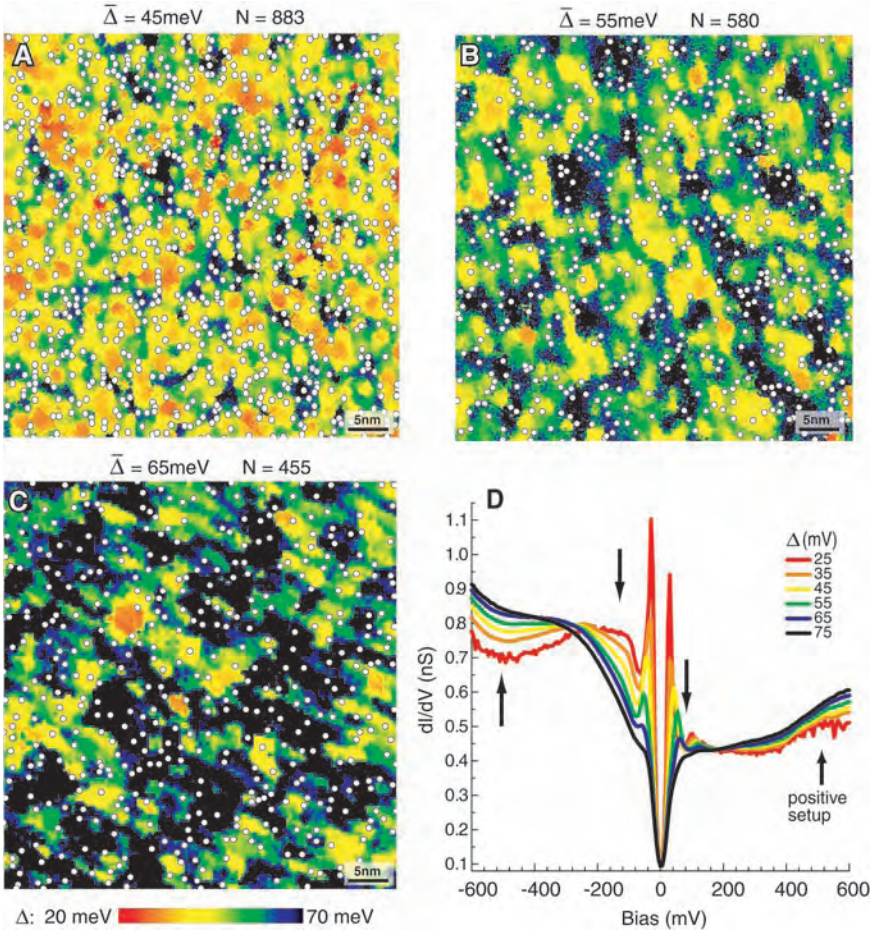


Fig. 2. (**A** to **C**) Three gap maps, with identical field of view sizes of 49 nm (corresponding to ~16,000 CuO_2 plaquettes) and color scales, measured on samples with three different oxygen doping levels. Average gap magnitudes $\bar{\Delta}$ are shown at the top, together with values of N, the total number of –0.96 V features (shown as white circles) in the field of view. Those in (**B**) were derived from the LDOS image in Fig. 1B. Note that the black regions have a distinct spectrum with a wide gap and no superconducting coherence peaks (*23*), as also shown by the black curve in (**D**). (**D**) The average spectrum associated with each superconducting gap value is color-coded to associate each gap-average spectrum with regions of the same color in all gap maps. As coherence peaks decrease and local gap values increase near dopant defects, the spectral weight shifts from below to above 250 meV for filled states and from below to above 150 meV for empty states. To make these observations, we used setup voltages of +600 meV and –600 meV, yielding simultaneous topographs with minimal nanoscale disorder and almost identical sets of spectra. Thus, no renormalization of these spectra to compensate for topographic disorder (*11*) was required. Further, the –600 meV setup voltage is sufficiently far below the –960 meV peak at the dopant sites to avoid tip elevation errors due to the impurity states.

How do the dopant defects generate the electronic disorder? Because disorder in topography clearly exists at low-bias $V_B \leq \pm 0.25$ V but diminishes markedly above $V_B \approx \pm 0.60$ V, the low-bias electronic disorder must be compensated by contributions of the opposite sign beginning above $V_B \approx \pm 0.25$ V. This can occur only if the low-energy LDOS spectral weight diminutions near dopant defects are balanced by equal enhancements at intermediate energies. To test this deduction, we measured LDOS$(\vec{r},E)$ images up to high voltage (but well below the ~ –0.96 eV impurity state) and

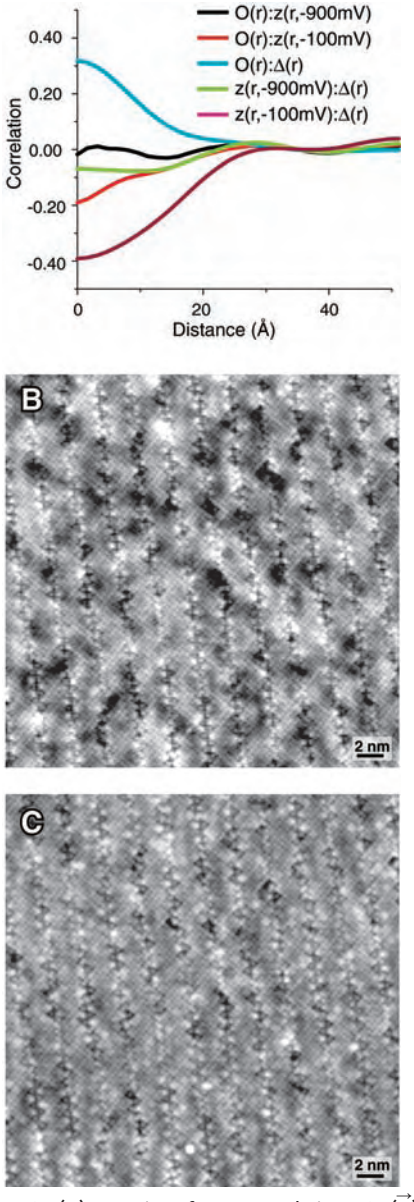


Fig. 3. (**A**) A series of cross-correlations $C_{f,g}(\vec{R})$ among the simultaneously measured $z(\vec{r},V_B)$, $\Delta(\vec{r})$, and $O(\vec{r})$ images. These show the angular average $C_{f,g}(|\vec{R}|)$ for *f*:*g* pairs $O(\vec{r})$:$\Delta(\vec{r})$ (blue line), $O(\vec{r})$:$z(\vec{r},V_B = -100$ meV) (red line), $O(\vec{r})$:$z(\vec{r},V_B = -900$ meV) (black line), $\Delta(\vec{r})$:$z(\vec{r},V_B = -100$ meV) (purple line), and $\Delta(\vec{r})$:$z(\vec{r},V_B = -900$ meV) (green line). We note that because $O(\vec{r})$ is an image with intensity in only a few percent of this field of view (see Fig. 1B), any correlation $C[O(\vec{r}), f(\vec{r})] \sim 0.3$ is actually quite strong. The images in Fig. 2, A to C, show directly that the $O(\vec{r})$:$\Delta(\vec{r})$ correlation, while strong, is not obviously deterministic, as might be expected for a nonlocal quantum mechanical system. (**B**) A typical low-bias topographic image $z(\vec{r},V_B = -100$ mV), with the 2.6 nm "supermodulation" surface undulation removed but retaining the crystalline structure, and showing the low-bias disorder as dark patches. (**C**) A high-bias topographic image $z(\vec{r},V_B = -900$ mV), from the same field of view as (B) with the same gray scale, with the 2.6 nm "supermodulation" surface undulation removed but retaining the crystalline structure. Here, minimal nanoscale electronic disorder can be detected.

identified the average spectrum associated with each superconducting gap value. The results (Fig. 2D) are color-coded to associate each gap-averaged spectrum with regions of the same color in all gap maps. These shifts are correlated with dopant defect locations to the same degree that the gap map is (Fig. 2, A to C; Fig. 3A, blue line). Therefore, as coherence peaks are suppressed and local gap values increase near dopant defects (Fig. 2, A to C; Fig. 3A), spectral weight shifts from below to above $E = 250$ meV for filled states and $E = 150$ meV for empty states are seen (arrows, Fig. 2D). Total spectral weight is conserved within the range $|E| < 0.9$ eV. These effects are most easily detected at intermediate doping ($p \approx 0.14$); we anticipate that they will diminish at very low and very high doping.

Because charge density variations are significantly weaker than previously suspected, a non–charge-driven mechanism for the nanoscale electronic disorder effects should be considered. It is known that substitutional atomic disorder at the Sr site suppresses superconductivity strongly (*4*), most likely by distortions of the crystalline unit cell and associated changes in the hopping matrix elements and superconducting correlations (*2*–*4*, *28*, *29*). Dopant atoms might act similarly, perhaps by displacing the Sr or apical oxygen atoms nearby, thereby changing the vibration frequencies and/or hopping integrals. Very recently, on the basis of the results reported here, Nunner *et al*. proposed a new model for the cuprate electronic disorder mechanism (*32*) in which the dopants not only produce some charge disorder, but also strongly perturb the local superconducting pairing potential on individual bonds between Cu sites. This proposal, that atomic-scale electron pairing–potential variations are paramount, offers a novel explanation for the complex phenomenology described here and in (*8*–*11*, *20*–*23*).

Correlating dopant defects with quasiparticle interference modulations. Another long-standing mystery has been the source of the very weak low-energy quasi-particle scattering interference modulations (*8*) and

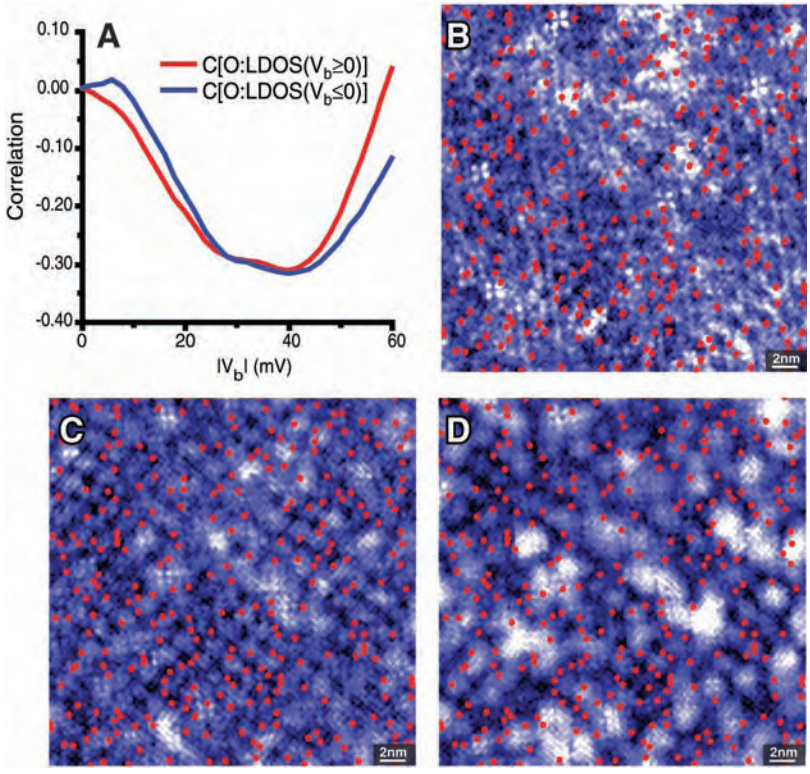


Fig. 4. (**A**) $R = 0$ cross-correlation values $C[O(r),LDOS(r,E)]$ for all energies $|E| < 60$ meV. These correlations begin to be significant around $E = 10$ meV and are intense for 30 meV $< |E| <$ 50 meV. The sign is always negative; hence, LDOS is low when the intensity of the –0.96 V peak in the $O(\vec{r})$ map (e.g., Fig. 1B) is high. This is true even though the LDOS modulation images look completely different at energies that differ by less than 10 meV. (**B**) Image of oxygen dopant defect locations (red dots) and LDOS at $E = -14$ meV. The modulations are weak at this energy and consist of numerous wave vectors. Nonetheless, the tendency for the dopant defects to lie in the minima of LDOS modulations can be detected by eye. (**C**) Image [simultaneous with (B) and (D)] of dopant defect locations and LDOS at $E = -24$ meV. Here the modulations are strong and more clearly reveal the tendency for the dopants to be at the minima. (**D**) Image [simultaneous with (B) and (C)] of dopant defect locations and LDOS at $E = -34$ meV. Here the long-wavelength modulations no longer exist, and the contrast basically represents the height of the –34 meV coherence peaks. The tendency for these peaks to be strongly diminished where the dopants are clustered is obvious.

the much stronger superconducting coherence peak suppressions (*8*, *10*, *22*, *23*) in Bi-2212. The analysis of correlations between the dopant defects and quasi-particle interference patterns $C[O(\vec{r}):LDOS(\vec{r},E)]$ shown in Fig. 4A reveals directly that they are strongly correlated over a wide energy range. The negative sign of these correlations means that minima of LDOS modulations preferentially occur at the dopant defects. This is true at all energies, even though the actual quasi-particle interference patterns are quite different (*8*). This can be seen directly in images of $LDOS(\vec{r},E)$ at $E = -14$ meV, -24 meV, and -34 meV (Fig. 4, B to D, respectively), with dopant defect locations shown as red dots. In all cases we see that the dopant defects have a very high probability of being found at the minima in the LDOS, even though the patterns are quite different between energies. This implies a fixing of the spatial phase of the LDOS modulations at scattering sites, an effect first seen in cuprates at oxygen vacancy defects in $YBa_2Cu_3O_{6-\delta}$ (*5*). Perhaps more important, Fig. 4D shows that suppression of superconducting coherence peaks is found primarily near the dopant defect clusters, and the bright regions with the sharp coherence peaks—usually associated with strong superconductivity—occur between them. The correspondence between dopant defect locations and LDOS minima for a variety of LDOS patterns occurring at different energies (Fig. 4) provides clear and direct evidence that these dopant defects generate the LDOS modulations and also suppress superconducting coherence peaks in Bi-2212.

Our simultaneous imaging of apparent dopant-induced impurity states and superconducting electronic structure points to solutions for several outstanding problems. The results provide direct evidence for the concept of an atomic-scale source for the nanoscale electronic disorder in cuprates (*11*, *24–26*). Strong correlations between dopant defect distributions and both gap map and coherence peak amplitude show that the dopant defects are responsible for most (but perhaps not all) of the superconducting electronic disorder in Bi-2212. Further, scattering leading to quasi-particle interference can now be ascribed almost completely to whichever atomic-scale perturbation produces the dopant defects. Finally, the topographic disorder and related superconducting electronic disorder are due empirically to spectral weight shifts from low to high energy near each dopant defect. These data indicate that high-energy spectral weight redistributions, strong coherence peak suppressions, and very weak scattering of low-energy quasi-particles are dominant elements in the atomic-scale mechanism of superconducting electronic disorder in $Bi_2Sr_2CaCu_2O_{8+\delta}$. Similar phenomena are likely to be common in all nonstoichiometric oxygen-doped high-T_c cuprates.

References and Notes

1. S. Maekawa *et al.*, *Physics of Transition Metal Oxides* (Springer-Verlag, Berlin, 2004).
2. R. J. Cava, *Nature* **394**, 126 (1998).
3. J. P. Attfield, A. L. Kharlanov, J. A. McAllister, *Nature* **394**, 157 (1998).
4. H. Eisaki *et al.*, *Phys. Rev. B* **69**, 064512 (2004).
5. H. L. Edwards *et al.*, *Phys. Rev. Lett.* **75**, 1387 (1995).
6. Y. Kohsaka *et al.*, *Phys. Rev. Lett.* **93**, 097004 (2004).
7. T. Hanaguri *et al.*, *Nature* **430**, 1001 (2004).
8. K. McElroy *et al.*, *Nature* **422**, 520 (2003).
9. G. Kinoda *et al.*, *Phys. Rev. B* **67**, 224509 (2003).
10. K. M. Lang *et al.*, *Nature* **415**, 412 (2002).
11. S. H. Pan *et al.*, *Nature* **413**, 282 (2001).
12. D. Grebille *et al.*, *Acta Crystallogr.* **B52**, 628 (1996).
13. W. Que, M. B. Walker, *Phys. Rev. B* **46**, 14772 (1992).
14. N. Miyakawa *et al.*, *Phys. Rev. Lett.* **80**, 157 (1998).
15. Q.-H. Wang, D.-H. Lee, *Phys. Rev. B* **67**, 020511 (2003).
16. L. Capriotti, D. J. Scalapino, R. D. Sedgewick, *Phys. Rev. B* **68**, 014508 (2003).
17. L. Zhu, W. A. Atkinson, P. J. Hirschfeld, *Phys. Rev. B* **69**, 060503 (2004).
18. U. Chatterjee *et al.*, available at http://arXiv.org/abs/cond-mat/0505296.
19. K. McElroy *et al.*, available at http://arXiv.org/abs/cond-mat/0505333.
20. T. Cren *et al.*, *Europhys. Lett.* **54**, 84 (2001).
21. A. Matsuda *et al.*, *Physica C* **388**, 207 (2003).
22. C. Howald, P. Fournier, A. Kapitulnik, *Phys. Rev. B* **64**, 100504 (2001).
23. K. McElroy *et al.*, *Phys. Rev. Lett.* **94**, 197005 (2005).
24. Z. Wang *et al.*, *Phys. Rev. B* **65**, 064509 (2002).
25. Q.-H. Wang, J. H. Han, D.-H. Lee, *Phys. Rev. B* **65**, 054501 (2002).
26. I. Martin, A. V. Balatsky, *Physica C* **357–360**, 46 (2001).
27. D. J. Scalapino, T. S. Nunner, P. J. Hirschfeld, in *Proceedings of the 7th International Conference on Spectroscopies in Novel Superconductors* (*Sitges, Spain*), in press (available at http://arXiv.org/abs/cond-mat/0409204).
28. J.-X. Zhu *et al.*, *Phys. Rev. Lett.* **91**, 057004 (2003).
29. E. Kaneshita, I. Martin, A. R. Bishop, *Phys. Soc. Jpn. Lett.* **73**, 3223 (2004).
30. J. Bobroff *et al.*, *Phys. Rev. Lett.* **89**, 157002 (2002).
31. J. W. Loram, J. L. Tallon, W. Y. Liang, *Phys. Rev. B* **69**, 060502 (2004).
32. T. Nunner *et al.*, available at http://arXiv.org/abs/cond-mat/0504693.

33. We thank W. Atkinson, A. V. Balatsky, A. Bishop, J. C. Campuzano, R. J. Cava, P. Hirschfeld, J. E. Hoffman, V. Madhavan, R. Markiewicz, T. M. Rice, G. Sawatzky, D. J. Scalapino, Z.-X. Shen, Z. Wang, and A. Yazdani for helpful conversations and communications. Supported by the Office of Naval Research, NSF, and Cornell University.

4 April 2005; accepted 30 June 2005
10.1126/science.1113095

FD, a bZIP Protein Mediating Signals from the Floral Pathway Integrator FT at the Shoot Apex

Mitsutomo Abe,[1]* Yasushi Kobayashi,[1,2]*† Sumiko Yamamoto,[1,2]* Yasufumi Daimon,[1] Ayako Yamaguchi,[1] Yoko Ikeda,[1] Harutaka Ichinoki,[1] Michitaka Notaguchi,[1] Koji Goto,[2,3] Takashi Araki[1,2,4]‡

***FLOWERING LOCUS T* (*FT*) is a conserved promoter of flowering that acts downstream of various regulatory pathways, including one that mediates photoperiodic induction through *CONSTANS* (*CO*), and is expressed in the vasculature of cotyledons and leaves. A bZIP transcription factor, FD, preferentially expressed in the shoot apex is required for FT to promote flowering. FD and FT are interdependent partners through protein interaction and act at the shoot apex to promote floral transition and to initiate floral development through transcriptional activation of a floral meristem identity gene, *APETALA1* (*AP1*). *FT* may represent a long-distance signal in flowering.**

Flowering in *Arabidopsis* is regulated by several pathways that converge on the transcriptional regulation of the floral pathway integrators *FT*, *SUPPRESSOR OF OVEREXPRESSION OF CO 1* (*SOC1*), and *LEAFY* (*LFY*) (*1*). *FT* is a direct target of CO, a key transcriptional regulator of the photoperiod pathway, and the role of *FT* as a potent promoter of flowering in response to photoperiods is conserved in *Arabidopsis* and rice (*2–6*). *FT* is expressed in the phloem tissues of cotyledons and leaves (*7*, *8*) and encodes a 20-kD protein with homology to phosphatidylethanolamine binding protein or Raf kinase inhibitor protein (*2*, *3*). However, the biochemical function of FT and downstream events leading to floral transition and floral morphogenesis at the shoot apex remain unknown.

bZIP protein FD is required for FT function. To understand how signals are mediated from *FT* to finally cause floral transition and floral morphogenesis, we searched for genes required for *FT* to promote flowering. Ectopic

[1]Department of Botany, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan. [2]CREST, Japan Science and Technology Agency, Kawaguchi 332-0012, Japan. [3]Research Institute for Biological Sciences Okayama, Okayama 716-1241, Japan. [4]Adjunct Division of Applied Genetics, National Institute of Genetics, Mishima 411-8540, Japan.

*These authors contributed equally to this work.
†Present address: Department of Molecular Biology, Max Planck Institute for Developmental Biology, D-72076 Tübingen, Germany.
‡To whom correspondence should be addressed. E-mail: taraqui@cosmos.bot.kyoto-u.ac.jp

Table 1. Flowering times of transgenic and mutant plants.

Genotype*	No. of rosette leaves†	No. of cauline leaves†	*N*
	Experiment 1, LDs		
Wild type (Col)	14.2 ± 2.2 (11–18)	3.0 ± 0.7 (2–4)	25
fd-1	23.1 ± 3.1 (19–29)	6.6 ± 1.2 (4–8)	14
35S::FT (YK#1-5C)	2.8 ± 0.4 (2–3)	1.0 ± 0.4 (0–2)	25
35S::FT (YK#1-5C); *fd-1*	18.4 ± 2.7 (14–24)	5.4 ± 1.2 (4–8)	20
soc1-101D‡	1.9 ± 0.3 (1–2)§	2.1 ± 0.3 (2–3)	15
soc1-101D; fd-1‡	2.0 ± 0 (2)	1.9 ± 0.3 (1–2)	21
Wild type (L*er*)	7.2 ± 1.1 (6–9)	2.4 ± 0.8 (1–4)	22
fd-1	12.3 ± 0.7 (11–15)	4.1 ± 0.7 (3–5)	43
35S::FT (YK#1-5L)	2.0 ± 0 (2)	1.2 ± 0.4 (1–2)	40
35S::FT (YK#1-5L); *fd-1*	4.0 ± 0 (4)	2.4 ± 0.8 (1–4)	24
	Experiment 2, SDs		
Wild type (Col),	45.3 ± 5.1 (38–54)	10.4 ± 2.0 (7–15)	17
fd-1	51.0 ± 5.8 (43–61)	9.8 ± 1.7 (7–13)	15
35S::FT (YK#1-5C)	2.3 ± 0.5 (2–3)	1.0 ± 0.5 (0–2)	25
35S::FT (YK#1-5C); *fd-1*	18.2 ± 4.3 (11–26)	4.4 ± 1.1 (3–7)	19
	Experiment 3, LDs		
Wild type (L*er*)	5.3 ± 0.5 (5–6)	1.7 ± 0.5 (1–2)	14
ft-3‡	11.2 ± 0.8 (10–12)	2.8 ± 0.4 (2–3)	10
FD::FT (YD#2-1); *ft-3*	3.5 ± 0.5 (3–4)	0.5 ± 0.5 (0–1)	14
PDF1::FT (MA#4-12); *ft-3*	2.1 ± 0.4 (2–3)	1.0 ± 0.5 (0–2)	35
SULTR2;1::FT (YD#1-1); *ft-3*	3.0 ± 0 (3)	0.7 ± 0.5 (0–1)	15
IAA14::FT (YD#4-3); *ft-3*‡	10.7 ± 0.8 (9–12)	2.4 ± 0.5 (2–3)	20
	Experiment 4, LDs		
Wild type (Col)	10.0 ± 0.7 (9–12)	2.4 ± 0.6 (1–3)	21
fd-1‡	21.1 ± 2.0 (18–24)	5.5 ± 0.5 (5–6)	22
FD::FD (MN#7-1); *fd-1*	8.4 ± 0.5 (8–9)	2.9 ± 0.5 (2–4)	21
SULTR2;1::FD (MN#4-2); *fd-1*‡	20.7 ± 1.6 (18–24)	5.0 ± 0.7 (4–6)	23
	Experiment 5, LDs		
Wild type (L*er*)	5.3 ± 0.5 (5–6)	1.8 ± 0.4 (1–2)	15
ft-3	13.7 ± 0.7 (12–15)	3.3 ± 0.7 (2–5)	15
ft-3; fd-1	17.8 ± 0.8 (16–19)	5.3 ± 0.4 (5–6)	16
FD::FT (YD#2-1); *ft-3*	2.9 ± 0.6 (2–4)	0.8 ± 0.4 (0–1)	15
FD::FT (YD#2-1); *ft-3; fd-1*	5.9 ± 0.6 (5–7)	1.8 ± 0.6 (1–3)	15
PDF1::FT (MA#4-12); *ft-3*	2.0 ± 0 (2)	1.3 ± 0.5 (1–2)	18
PDF1::FT (MA#4-12); *ft-3*; *fd-1*	6.4 ± 0.6 (5–7)	2.5 ± 0.5 (2–3)	20
SULTR2;1::FT (YD#1-1); *ft-3*	2.7 ± 0.5 (2–3)	1.2 ± 0.4 (1–2)	13
SULTR2;1::FT (YD#1-1); *ft-3; fd-1*	4.1 ± 0.3 (4–5)	2.0 ± 0 (2)	13
	Experiment 6, LDs		
Wild type (L*er*), –Dex	6.8 ± 0.9 (5–8)	3.0 ± 0.6 (2–4)	22
Wild type (L*er*), +Dex‖	6.4 ± 0.9 (5–8)	2.9 ± 0.7 (2–4)	24
ft-3, −Dex	14.1 ± 2.5 (11–19)	4.9 ± 1.0 (4–7)	12
ft-3, +Dex‖	12.7 ± 1.8 (10–15)	4.2 ± 0.6 (3–5)	11
35S::FT:GR (YD#9-a); *ft-3*, −Dex	8.8 ± 1.7 (6–12)	3.7 ± 0.6 (3–5)	24
35S::FT:GR (YD#9-a); *ft-3*, +Dex	3.3 ± 0.7 (2–4)	2.1 ± 0.4 (1–3)	24

*Genetic background: Col, Columbia; L*er*, Landsberg *er*. Plants in each experiment were grown under LDs or SDs as indicated. +Dex indicates Dex treatment (*14*). †Indicators of flowering time (*10*) and shown as average ± SD (range). Statistical tests were done on the number of rosette leaves. ‡No statistically significant difference among indicated genotypes in each experiment (Student's *t* test, $P > 0.1$). No symbol means that there was a statistically significant difference ($P < 0.001$) among the genotypes or conditions compared in each experiment. §Includes two plants with elongated internodes. ‖Indicates no statistically significant difference from the above condition ($P > 0.06$).

overexpression of *FT* by the 35*S* RNA promoter of cauliflower mosaic virus (*35S::FT*) causes a precocious-flowering phenotype (*2*, *3*). We screened for suppressors of *35S::FT*, because a similar approach with *35S::CO* was successful in elucidating the downstream targets of CO, *FT* and *SOC1* (*4*, *9*). In addition to screening mutagenized populations of *35S::FT*, we examined known late-flowering mutants (*10*) for their effect on the *35S::FT* phenotype. Through the latter approach, we found that *fd-1* is a strong suppressor of *35S::FT* (Table 1). In contrast to a strong effect on *35S::FT*, *fd-1* had only a weak effect on a similar precocious-flowering phenotype of *soc1-101D*, an activation tagged allele of *SOC1* (*11*) (Table 1), or *35S::LFY* (*12*). These observations suggest that the *FD* activity is required specifically for the promotion of flowering by *FT*.

The *FD* gene was identified with *AtbZIP14* (At4g35900) (*13*) by a map-based approach (figs. S1 and S2 and table S1). In seedlings, *FD* expression was observed mainly in the shoot apex (Fig. 1, A to E and J), did not show distinct circadian oscillation, and was not affected by photoperiods and *CO* activity (fig. S3). Under both short days (SDs) and long days (LDs), *FD* mRNA levels increased with time after germination (fig. S3). To examine the subcellular distribution of FD protein, we made a construct

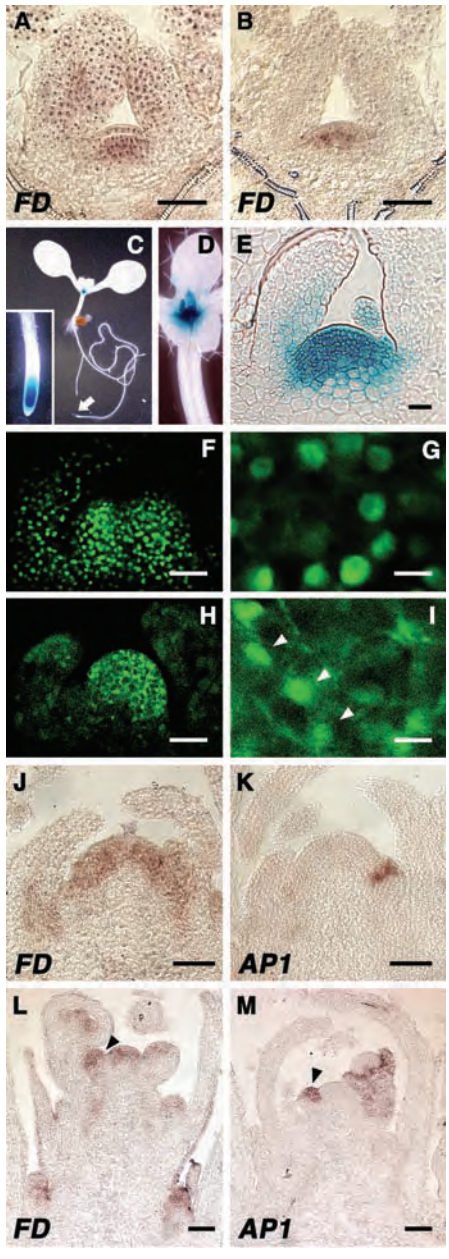


Fig. 1. Expression of *FD* and subcellular distribution of EGFP:FD and FT:EGFP. (**A** and **B**) Expression of *FD* in shoot apex of wild-type seedlings grown under LDs (16 hours light) for 6 (A) and 7 (B) days. (**C** to **E**) GUS staining of *gFD::GUS* seedlings grown under LDs for 6 (C and D) and 7 (E) days. (C) A whole seedling. The arrow indicates a root tip, which is enlarged in the inset. (D) Shoot apical region. (E) Longitudinal section of shoot apex. (**F** to **I**) Distribution of functional EGFP:FD fusion protein (F and G) and functional FT:EGFP fusion protein (H and I) expressed in shoot apex by *FD* promoter. (G) and (I) are enlargements showing subcellular distribution. Arrowheads in (I) indicate nuclei. (**J** to **M**) Expression of *FD* (J and L) and *AP1* (K and M) in shoot apex of wild-type plants at floral transition on day 10 (J and K) and early inflorescence stage on day 15 (L and M) under LDs. Arrowheads in (L) and (M) indicate floral meristems at stage 1. Scale bars: 50 μm in (F), (H), (L), and (M); 20 μm in (A), (B), (E), (J), and (K); and 10 μm in (G) and (I).

to express enhanced green fluorescent protein–FD (EGFP:FD) fusion protein by the *FD* promoter (*FD::EGFP:FD*) (*14*). That the fusion protein is functional was confirmed by the rescue of *fd-1* (table S2). EGFP:FD was localized in the nucleus in cells of the shoot apex (Fig. 1, F and G). Similarly, nuclear localization of functional enhanced yellow fluorescent protein–FD (EYFP:FD) fusion protein (table S2) was observed in cells of the shoot apex and other tissues in *35S::EYFP:FD* seedlings with various genetic backgrounds (fig. S4) (*14*), suggesting a constitutive nuclear localization.

To gain clues to the molecular basis of the requirement of *FD* for *FT* function, we investigated protein interactions. FT interacted with FD in yeast cells (Fig. 2A) and in vitro (Fig. 2C). In contrast, FD showed very weak interaction in yeast cells with TERMINAL FLOWER 1 (TFL1), the FT-related protein with an antagonistic role in the regulation of flowering (*2*, *3*) (Fig. 2A and fig. S5). Because the subcellular distribution of the FT protein remains unknown, we examined the distribution of functional FT:EGFP fusion protein (table S2) expressed in the shoot apex of *Arabidopsis* by the *FD* promoter or in leaf epidermal cells of *Nicotiana benthamiana* by the 35*S* promoter (*14*). In both cases, FT:EGFP was observed in the nucleus and cytoplasm (Fig. 1, H and I, and fig. S4). That FT is able to function in the nucleus was supported by observations that FT protein fused to a glucocorticoid receptor (GR) expressed by the 35*S* promoter (*35S::FT:GR*) (*14*) promoted flowering on dexamethasone (Dex) treatment (Table 1). These findings suggest that FD and FT proteins coexist in the nucleus. We further analyzed the interaction of FT and FD proteins in plant cells using bimolecular fluorescent complementation (BiFC) (*15*). In tobacco leaf epidermal cells coexpressing the N-terminal half of EYFP fused to FD (YN-FD) and the C-terminal half of EYFP fused to FT (YC-FT) (*14*), YFP fluorescence was observed in the nucleus (Fig. 2D). These findings show that protein interaction is the basis of the dependence of FT on FD.

Floral meristem identity genes are regulatory targets of FT and FD. We next tried to identify the regulatory targets of FT and FD in the shoot apex that cause floral transition and morphogenesis. We obtained clues from analysis of *ft; lfy* and *fd; lfy* double mutants. A previous work showed that *ft; lfy* greatly reduced mRNA levels of *AP1* and caused severe defects in floral meristem specification (*16*). These observations led to the suggestion that *FT* and *LFY* play redundant roles in up-regulation of *AP1* and floral fate specification (*1*, *16*). In agreement with a previous report, *ft; lfy* greatly reduced the amount and region of *AP1* expression in the shoot apex and caused severe defects in floral development (Fig. 3, A to D, and fig. S6). *fd; lfy* plants had an inflorescence phenotype indistinguishable from that of *ft; lfy* and displayed a severely reduced amount and spatial extent of *AP1* expression (Fig. 3, A to D, and fig. S6). These observations suggest that *FT* and *FD* together are involved in the regulation of *AP1* redundantly with *LFY*.

Consistent with a role of FD in the regulation of *AP1*, ectopic expression of *AP1* was observed in *35S::FD* seedlings (Fig. 3E). Ectopic induction of *AP1* expression by FD was abolished by *ft* mutation (Fig. 3F) or under SDs, which reduce *FT* expression (*2*, *3*, *17*, *18*) (Fig. 3G). Upon a shift from SDs to LDs, which induces *FT* expression (*8*, *18*), *AP1* expression was induced (Fig. 3H). In young seedlings, *FT* is expressed mainly in the vasculature of cotyledons, and little is detected in mesophyll cells (*7*, *8*). In *35S::FD* seedlings, *AP1* expression was detected in the vascular-rich fraction but not in the mesophyll-rich fraction from cotyledons (Fig. 3I). These observations suggest that activation of *AP1* expression by FD requires the *FT* function. Finally, in the shoot apex around the stage of floral transition and in the young inflorescence apex, the region of *AP1* expression was within the expression domain of *FD* (Fig. 1, J to M). Thus, *AP1* seems to be a regulatory target of FD, which requires the FT activity through protein interaction. In support of this conclusion, several potential bZIP protein binding motifs were found in a 1.7-kb *AP1* promoter (*19*) (fig. S7).

Because *AP1* expression was observed only in a subset of the expression domain of *FD*, there should be factors that restrict *AP1* expression to nascent lateral meristems. *TFL1*, which has a role antagonistic to *FT* (*2*, *3*), is likely to be responsible for suppressing *AP1* expression in the shoot apical meristem proper (*20*, *21*). That the loss of *AP1* alone does not affect the precocious-flowering phenotype of *35S::FT* (*3*) suggests that other regulatory targets of FD contribute to the promotion of floral transition. *FRUITFULL* (*FUL*) and *CAULIFLOWER* (*CAL*), which act redundantly with *AP1* to promote flowering (*22*), are candidates for such targets (Fig. 3, C and E to J, and fig. S8).

How FT regulates FD activity is another important question. Constitutive nuclear localization of FD and the presence of FT in the nucleus suggest the regulation of FD activity in the nucleus. Whether FD and FT form a stable transcriptional complex or interact only transiently remains to be investigated. FD protein has a potential phosphorylation site for Ca^{2+}-dependent protein kinases (CDPKs) at the C terminus (Fig. 2B and fig. S2). Deletion or mutation of this

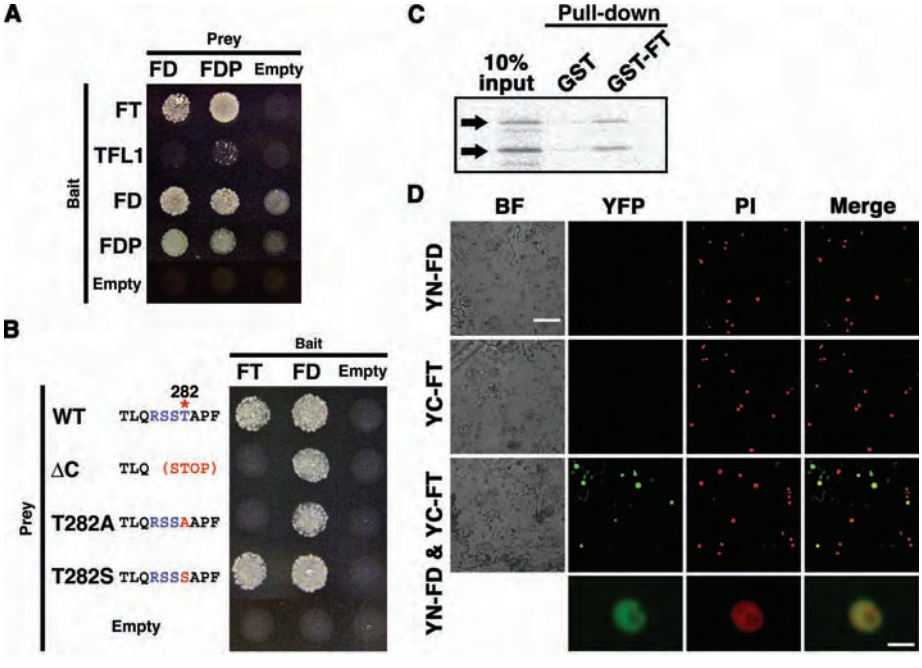


Fig. 2. Protein interaction. (**A**) Yeast two-hybrid assay of interaction among FD, FDP, FT, and TFL1. FDP, FD PARALOG (AtbZIP27). (**B**) Yeast two-hybrid assay of interaction between FT and C-terminal mutants of FD. C-terminal sequences of wild-type (WT) FD and mutant (ΔC, T282A, T282S) FD proteins including a possible CDPK site (blue). The asterisk on the WT sequence indicates a threonine (T) residue expected to be phosphorylated. T282A (substitution with an alanine), but not T282S (substitution with a serine), abolishes the CDPK site. None of the mutants affected interaction with wild-type FD. (**C**) In vitro pull-down assay of interaction between FT and FD (*14*). "10% input" indicates 10% of ^{35}S-labeled FD subjected to pull-down by GST or GST-FT. Arrows indicate labeled FD. (**D**) BiFC analysis of interaction between FT and FD in *N. benthamiana* leaf epidermis (*14*). BF, blight field image; YFP, YFP fluorescence; PI, propidium iodide fluorescence (nuclei); Merge, merge of YFP and PI; YN-FD, expression of YN-FD alone; YC-FT, expression of YC-FT alone; YN-FD & YC-FT, coexpression of YN-FD and YC-FT. At the bottom are higher magnification images of a nucleus of a cell coexpressing YN-FD and YC-FT. Not all nuclei were stained with PI. Scale bars: 100 μm (upper three rows) and 10 μm (bottom row).

site abolished interaction with FT (Fig. 2B), ectopic induction of *AP1* expression (Fig. 3J), and the ability to complement *fd-1* (fig. S9), although nuclear localization was not affected (fig. S4). These findings suggest the importance of phosphorylation of FD in the interaction with FT and in its functional regulation.

Mutual dependence and site of action of FT and FD. The mutual dependence of *FT* and *FD*, as shown above, is further supported by the observation that flowering of *35S::EYFP:FD* plants was delayed under SDs, which reduce *FT* expression (table S3). Furthermore, the enhanced phenotype in *35S::FT; 35S::FD* (tables S3 and S4) indicates that *FT* and *FD* are mutually limiting for the combined activity of *FT* and *FD*. These raise the question of the site(s) of action of FT and FD. In seedlings, *FT* is expressed in the vasculature of cotyledons, but not in the shoot apex (*7*) (fig. S10), whereas *FD* is expressed in the shoot apex but not in cotyledons and leaves (Fig. 1, A to E, and fig. S10). As expected, restoration of the *FT* function in the vasculature through expression by *SULTR2;1::FT* (*14*) could rescue the late-flowering phenotype of *ft* (Table 1 and table S5). Restoration of the *FT* function in the shoot apex of *ft*, either in the whole region by *FD::FT* or in the outermost cell layer (L1) by *PDF1::FT* (*14*), also rescued the late-flowering phenotype (Table 1 and table S5). These findings agree well with those of a previous report that the late-flowering phenotype of *co* is suppressed by similar constructs for *FT* misexpression (*23*). We further observed that *FD::FT*, *PDF1::FT*, and *SULTR2;1::FT* rescued delayed flowering, reduced *AP1* expression, and severe floral defect in *ft; lfy* (Fig. 3, B and C, and fig. S6). By contrast, restoration of *FT* in root vasculature by *IAA14::FT* (*14*) failed to rescue *ft* (Table 1). These observations indicate that ectopically expressed *FT* in the shoot apex can exert an effect on flowering. *FT* expressed in the shoot apex requires the *FD* function, because *fd-1* attenuated the rescued phenotype (Table 1 and table S5). In contrast, *FD* rescued the late-flowering phenotype of *fd* through expression in the whole shoot apex (by *FD::FD*), but not through ectopic expression in leaf vasculature (by *SULTR2;1::FD*), where *FT* is expressed (Table 1). Therefore, *FD* acts in the shoot apex and seems to be required in all cell layers. These findings, together with observations that protein interaction is the basis for interdependence between FT and FD, suggest that the shoot apex is the site of the *FT* and *FD* action.

FT and the long-distance signal in flowering. It has long been believed that a long-distance signal, named florigen (*24*), is generated in leaves upon exposure to inductive photoperiods, is transported to the shoot apex, and acts there to promote flowering (*25*). However, the nature of the signal has remained elusive (*25*). Our present work supports an emerging hypothesis (*7*, *8*, *23*) that the *FT* products represent a part of the long-distance signal(s) generated in cotyledons and leaves (mainly in the phloem tissues) and act at the shoot apex to promote floral transition and to initiate floral development (fig. S11).

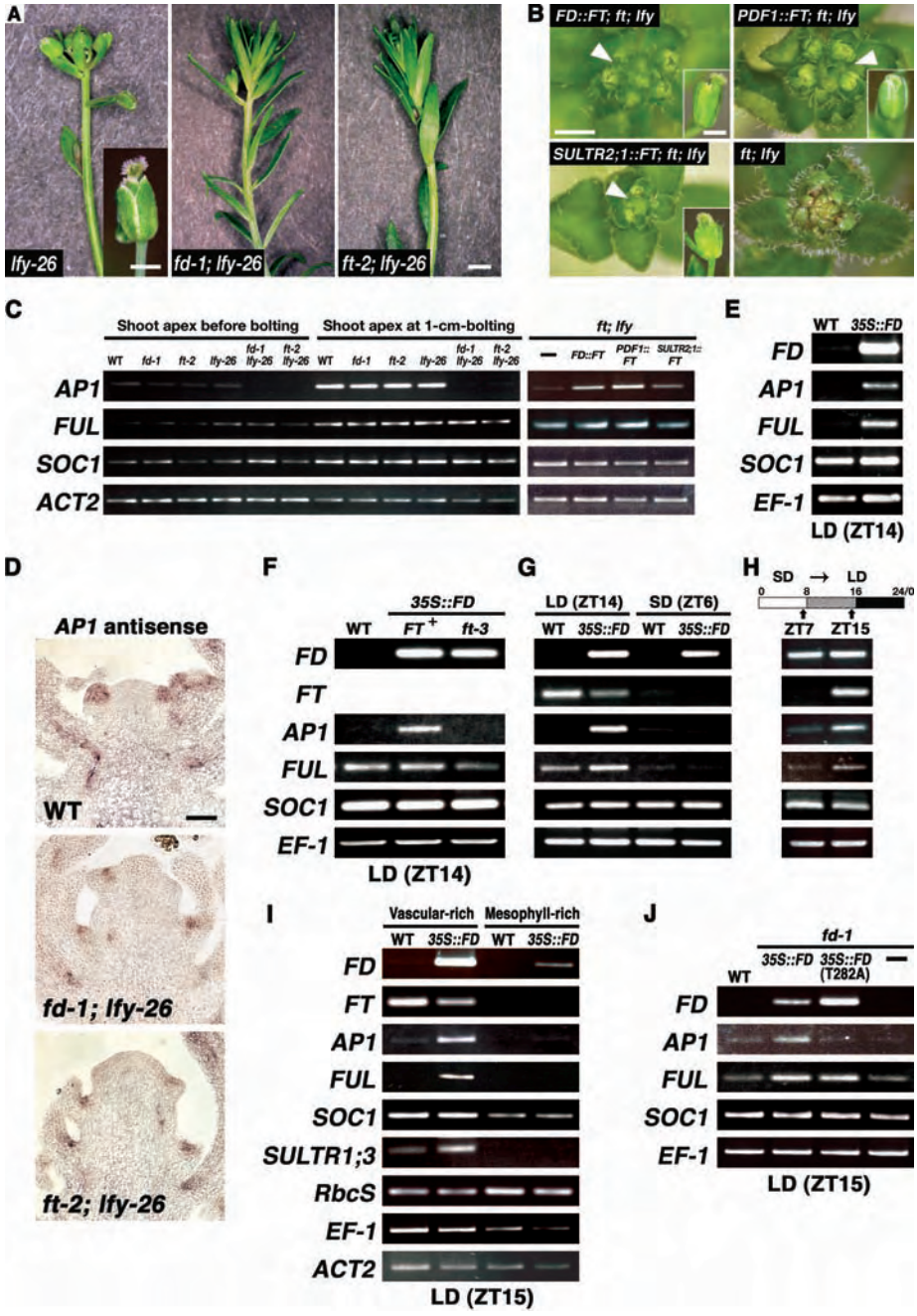


Fig. 3. *FT*-dependent activation of *AP1* by FD. (**A**) Floral defect in *lfy-26*, *fd-1; lfy-26*, and *ft-2; lfy-26*. Lateral structures on the primary inflorescence that would be single flowers in the wild-type plant. Inset shows a flower formed later in *lfy-26*. Scale bars: 2 mm and 1 mm (inset). (**B**) Rescue of floral defect in *ft-2; lfy-26* by tissue-specific expression of *FT*. Young primary inflorescences and flowers of *ft-2; lfy-26* with indicated *FT* constructs. Arrowheads indicate a terminal flower. Scale bars, 2 mm and 1 mm (inset). (**C**) *AP1* and *FUL* expression in shoot apex of various genotypes at two different stages. *SOC1* and *ACT2* were amplified for reference. (**D**) *AP1* expression in young inflorescence apex. Scale bar: 10 μm. (**E** to **J**) *AP1* and *FUL* expression in *35S::FD* and wild-type (WT) seedlings. *SOC1*, *EF-1*, and *ACT2* were amplified for reference. Whole seedlings (E to H and J) or cotyledons (I) were harvested for RNA extraction at the indicated Zeitgeber time (ZT) points. (E) Seven-day-old seedlings under LDs. (F) Seven-day-old *35S::FD* in FT^{+} and *ft-3* background and wild type under LDs. (G) Seven-day-old seedlings under LDs and SDs (8 hours light). (H) *35S::FD* seedlings grown for 6 days under SDs and subjected to day-length extension from 8 to 16 hours on day 7. (I) Vascular- and mesophyll-rich fractions from cotyledons of 10-day-old seedlings under LDs (*14*). *SULTR1;3* is a vascular marker and *RbcS* is preferentially expressed in mesophylls. (J) Seven-day-old seedlings of wild-type and *fd-1* with *35S::FD* or $35S::FD^{T282A}$ or without a transgene (−) under LDs.

References and Notes

1. G. G. Simpson, C. Dean, *Science* **296**, 285 (2002).
2. Y. Kobayashi, H. Kaya, K. Goto, M. Iwabuchi, T. Araki, *Science* **286**, 1960 (1999).
3. I. Kardailsky *et al.*, *Science* **286**, 1962 (1999).
4. A. Samach *et al.*, *Science* **288**, 1613 (2000).
5. S. Kojima *et al.*, *Plant Cell Physiol.* **43**, 1096 (2002).
6. R. Hayama, S. Yokoi, S. Tamaki, M. Yano, K. Shimamoto, *Nature* **422**, 719 (2003).
7. S. Takada, K. Goto, *Plant Cell* **15**, 2856 (2003).
8. A. Yamaguchi, Y. Kobayashi, K. Goto, M. Abe, T. Araki, *Plant Cell Physiol.* **46**, 1175 (2005).
9. H. Onouchi, M. I. Igeño, C. Périlleux, K. Graves, G. Coupland, *Plant Cell* **12**, 885 (2000).
10. M. Koornneef, C. J. Hanhart, J. H. van der Veen, *Mol. Gen. Genet.* **229**, 57 (1991).
11. H. Lee *et al.*, *Genes Dev.* **14**, 2366 (2000).
12. O. Nilsson, I. Lee, M. A. Blázquez, D. Weigel, *Genetics* **150**, 403 (1998).
13. M. Jakoby *et al.*, *Trends Plant Sci.* **7**, 106 (2002).
14. Materials and methods are available as supporting material on *Science* Online.
15. C. D. Hu, Y. Chinenov, T. K. Kerppola, *Mol. Cell* **9**, 789 (2002).
16. L. Ruiz-García *et al.*, *Plant Cell* **9**, 1921 (1997).
17. P. Suárez-López *et al.*, *Nature* **410**, 1116 (2001).
18. M. J. Yanovsky, S. A. Kay, *Nature* **419**, 308 (2002).
19. F. D. Hempel *et al.*, *Development* **124**, 3845 (1997).
20. O. J. Ratcliffe, D. J. Bradley, E. S. Coen, *Development* **126**, 1109 (1999).
21. S. J. Liljegren, C. Gustafson-Brown, A. Pinyopich, G. S. Ditta, M. F. Yanofsky, *Plant Cell* **11**, 1007 (1999).
22. C. Ferrándiz, Q. Gu, R. Martienssen, M. F. Yanofsky, *Development* **127**, 725 (2000).
23. H. An *et al.*, *Development* **131**, 3615 (2004).
24. M. K. Chailakhyan, *C. R. (Dokl.) Acad. Sci. URSS* **16**, 227 (1937).
25. J. A. D. Zeevaart, *Annu. Rev. Plant Physiol.* **27**, 321 (1976).
26. We thank P. Wigge and D. Weigel for the exchange of unpublished results; M. Koornneef, I. Lee, Y. Komeda, T. Takahashi, H. Takahashi, H. Fukaki, E. Lifschitz, G. Coupland and Plant Bioscience Limited, I. Hara-Nishimura, and the NSF-supported Arabidopsis Biological Resource Center for materials; M. Endo for technical advice; and Y. Tomita for technical assistance. Supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan (to T.A. and M.A.), the Core Research for Evolutional Science and Technology (CREST) program of the Japan Science and Technology Agency (to T.A.), and the Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN) program of the Bio-oriented Technology Research Advancement Institution, Japan (to M.A.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/309/5737/1052/DC1
Materials and Methods
SOM Text
Tables S1 to S5
Figs. S1 to S11
References

10 June 2005; accepted 12 July 2005
10.1126/science.1115983

Integration of Spatial and Temporal Information During Floral Induction in *Arabidopsis*

Philip A. Wigge,[1,4]*†‡ Min Chul Kim,[1]* Katja E. Jaeger,[1]† Wolfgang Busch,[2] Markus Schmid,[3] Jan U. Lohmann,[2] Detlef Weigel[1,4]‡

Flowering of *Arabidopsis* is regulated by several environmental and endogenous signals. An important integrator of these inputs is the *FLOWERING LOCUS T* (*FT*) gene, which encodes a small, possibly mobile protein. A primary response to floral induction is the activation of *FT* RNA expression in leaves. Because flowers form at a distant site, the shoot apex, these data suggest that *FT* primarily controls the timing of flowering. Integration of temporal and spatial information is mediated in part by the bZIP transcription factor FD, which is already expressed at the shoot apex before floral induction. A complex of FT and FD proteins in turn can activate floral identity genes such as *APETALA1* (*AP1*).

One of the major flowering pathways in *Arabidopsis*, the photoperiod pathway, positively regulates activity of the nuclear protein CONSTANS (CO), which acts upstream of a graft-transmissible signal produced in leaves (*1*–*3*). Experiments with a dexamethasone-dependent, constitutively expressed version of CO have suggested that CO directly activates genes with diverse biochemical functions (*4*). These include two genes that are known to promote flowering: the transcription factor gene *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*) and *FLOWERING LOCUS T* (*FT*), which encodes a small globular protein related to the floral repressor TERMINAL FLOWER 1 (TFL1) (*5*–*7*). In addition, *ACS10* and *P5CS2*, structural genes for ethylene and proline biosynthetic enzymes, were identified as potential CO targets in these experiments (*4*).

***FT* is the major primary target of CO in leaves.** Because the results with *CO* gain-of-function alleles had suggested that the CO-induced signal is complex, we used microarray analyses to identify targets of endogenous CO in leaves. To induce endogenous CO activity, we activated CO protein with light (*1*). We grew all plants in 8-hour short days. On the day of the experiment, we exposed the experimental group to 16 hours of light and the control group to 8 hours of light followed by 8 hours of darkness. We harvested leaves at the end of the 16-hour period. Differentially expressed genes were identified with a combination of per-gene variance [logit t test, $P < 0.025$ (*8*)] and common variance (>1.5× change).

Of 2000 genes that are activated or repressed after exposure to a single long day, merely three genes are responsive to long days in wild-type plants and at the same time differentially expressed between wild-type plants and *co* mutants in long days. Of these, only one gene does not respond at all to long days in a *co* background: *FT* (Fig. 1, A and B). In contrast, *SOC1*, *ACS10*, and *P5CS2* expression is independent of CO after exposure to a single long day (Fig. 1B), suggesting either that these genes respond only to higher levels of *CO* or that they respond to *CO* in tissues other than the leaf. In plants grown in continuous light, *FT* expression is much higher in leaves than at the shoot apex (Fig. 1C), consistent with leaves being the primary site of *FT* activation.

The finding that *FT* is the major target of *CO* in leaves is in agreement with our observation, also based on global expression profiles, that *FT* is the major output of *CO* at the shoot apex and that early floral markers such as *SOC1* and *FRUITFULL* (*FUL*) are similarly dependent on both *CO* and *FT* (*9*). Thus, the initial signal acting downstream of *CO* in leaves might be less complex than previously thought.

The bZIP protein FD is required for FT activity. To understand how FT activity is transduced, we searched for FT interactors. In a yeast two-hybrid screen, we isolated two closely related bZIP transcription factors, At2g17770 and At4g35900 (fig. S1), the *Arabidopsis* orthologs of tomato SIP8/SPGB, which interacts with *Arabidopsis* FT and TFL1, as well as the tomato TFL1 homolog SELF-PRUNING (*10*). The available collections of transferred DNA (T-DNA) insertion lines do not contain any At2g17770 alleles, but there are four different alleles but of At4g35900 (*11*). The late-flowering phenotype of these lines (Fig. 2A and Table 1) is rescued by a minigene, indicating that At4g35900 promotes flowering (fig. S2). The only late-flowering mutant described for this region of the genome is *fd*-1 (*12*), and complementation crosses showed *FD* and At4g35900 to be allelic. We therefore designated our reference insertion allele of At4g35900 as *fd*-2. By their phenotype and genetic interactions, *FD* and *FT* have been placed

[1]Department of Molecular Biology, [2]Arbeitsgruppe Lohmann, [3]Arbeitsgruppe Schmid, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany. [4]Salk Institute for Biological Studies, Plant Biology Laboratory, La Jolla, CA 92037, USA.

*These authors contributed equally to this work.
†Present address: John Innes Centre, Cell and Developmental Biology Department, Norwich NR4 7UH, UK.
‡To whom correspondence should be addressed. E-mail: philip.wigge@bbsrc.ac.uk (P.W.); weigel@weigelworld.org (D.W.)

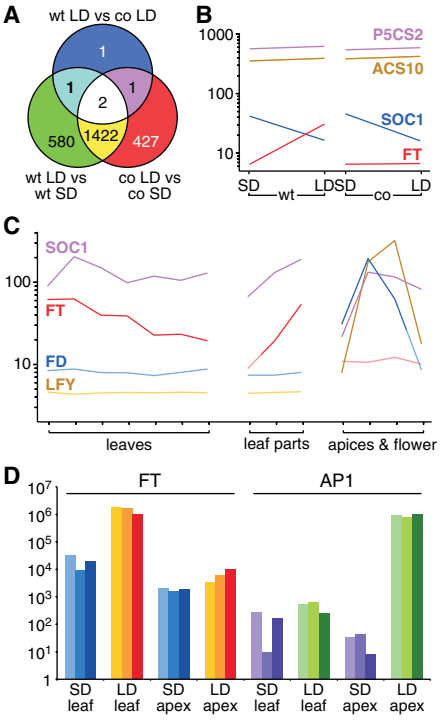


Fig. 1. Expression analyses using microarrays and real-time RT-PCR. (**A**) Venn diagram of Affymetrix ATH1 microarray data (ArrayExpress experiment E-TABM-2) showing that only a single gene, *FT*, is differentially affected by the *co* mutation and the shift from short days (SD) to long days (LD), indicated in bold. (**B**) Microarray values (averages of three replicates) for *FT* and three other genes previously identified as potentially direct CO targets (*4*). (**C**) Expression patterns of floral regulators, from the AtGenExpress microarray atlas (*26*). From left to right, tissues are embryonic leaves (cotyledons, 7 days) and rosette leaves 2, 4, 6, 8, 10, and 12 (17 days); petiole, proximal, and distal parts of leaf 7 (3 weeks); vegetative apex (7 days), transition apex (14 days), flowering apex (21 days), and stage-9 flowers. *FT* is expressed most strongly in the oldest leaves, and within a leaf, in the distal, most differentiated part. *SOC1* is expressed more strongly in leaves than in vegetative apices, but induced upon flowering at the shoot apex. Light color indicates an "absence" call by the Affymetrix MAS algorithm in all three replicates. (**D**) Real-time RT-PCR analysis of *FT* and *AP1* in leaf 3 and shoot apex of 20-day-old wild-type plants.

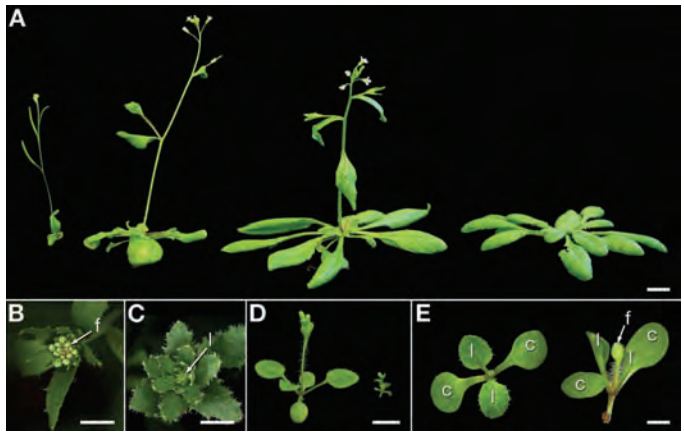


Fig. 2. Phenotypes of plants grown in long days. (**A**) Partial suppression of 35S:*FT* by *fd*-2. From left: 30-day-old 35S:*FT*, 35S:*FT* *fd*-2, wild-type, and *fd*-2 plants. (**B**) *lfy*-12 mutants produce abnormal flowers (f). (**C**) *lfy*-12 *fd*-2 plants produce leaves (l) instead. (**D**) An 18-day-old 35S:*FT* (left) compared with 35S:*FD* 35S:*FT* plant. (**E**) Cotyledons (c) and two normal leaves (l) are seen in a 10-day-old 35S:*FT tfl1 ft* plant (left). The first two leaves of 35S:*FT-VP16 tfl1 ft* are curled up and surround a single terminal flower (right). Plants in each panel were grown simultaneously. Scale bars, 1 cm (A), 5 mm [(B) to (D)], and 1 mm (E).

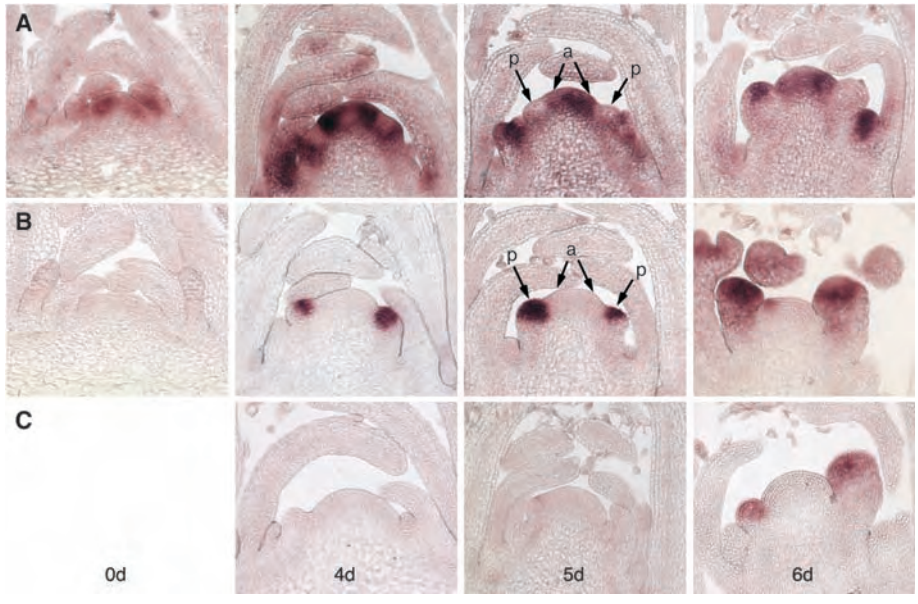


Fig. 3. In situ hybridization patterns of *FD* and *AP1* at the shoot apex. Plants transferred from short to long days were sampled at the end of each day. "0d" indicates last short day. A complete series until day 7 after transfer is shown in fig. S3. Day 4 after transfer ("4d") was the first day on which a robust *AP1* signal was apparent in the wild-type plant. (**A**) *FD* expression at the shoot apex of wild-type plants. Note transient expression in floral anlagen (a), with rapidly declining expression in floral primordia (p). (**B**) *AP1* expression in wild-type plants. *AP1* is detected in floral primordia and subsequent stages. (**C**) Delayed and weaker activation of *AP1* expression in *fd*-2 plants. A range of floral stages with corresponding variation in *AP1* expression was found in each sample. An average example is shown for each day and genotype.

in the same class of floral regulators (*12*, *13*). A special property of *ft* compared with several other late-flowering mutants is that *ft leafy* (*lfy*) double mutants never form flowers, indicating partially redundant functions of *FT* and *LFY* (*14*). *fd*-2 *lfy*-12 double mutants also lack flowers (Fig. 2, B and C). One of the outputs of combined *LFY* and *FT* activities is activation of the floral identity gene *AP1*. Again consistent with a role of *FD* in the *FT* pathway, activation of *AP1* is delayed in *fd*-2 mutants (Fig. 3, B and C). Finally, *FD* is required for *FT* activity, such that the early flowering caused by *FT* overexpression is partially suppressed by *fd*-2 (Fig. 2A and Table 1).

In contrast to *FT*, *FD* is highly expressed at the shoot apex (Fig. 1C). Specifically, *FD* RNA was observed in leaf and floral anlagen. Its levels decrease soon after floral primordia start to express *AP1* (Fig. 3, A and B, and fig. S3). Expression in floral anlagen appears to be slightly higher than in leaf anlagen, and a modest increase after floral induction is also seen with microarrays, both in wild-type plants and *ft* mutants (fig. S4).

It thus appears that *FT*, which is induced most strongly in leaves, integrates environmental signals that allow for correct timing of flower initiation, whereas *FD*, which is most strongly expressed at the shoot apex, provides spatial specificity for *FT* action. Because *FD* is already expressed at the shoot apex before floral induction, *FD* should be less limiting for flowering than *FT*. Indeed, the early flowering of lines that overexpress *FD* is weaker than that of 35S:*FT* plants (Table 1). In addition to early flowering, 35S:*FD* plants are small with curled leaves (Fig. 2D). Both phenotypes largely disappear in short days, when little *FT* activity is present (Table 1). That *FT* and *FD* are mutually limiting is further supported by the synergistic phenotype of 35S:*FT* 35S:*FD* plants (Fig. 2D).

***FT* and *FD* are sufficient to activate the expression of floral marker genes.** The strong leaf phenotype of 35S:*FD* plants, which

is reminiscent of plants that overexpress *AP1* or *FUL* (*4*, *15*), suggested that combined *FT* and *FD* activities are largely sufficient for triggering the floral transcriptional program. Consistent with this hypothesis, *AP1* and *FUL* are ectopically activated in the leaves of 35S:*FD* plants, but only during long days (Fig. 4, A and B).

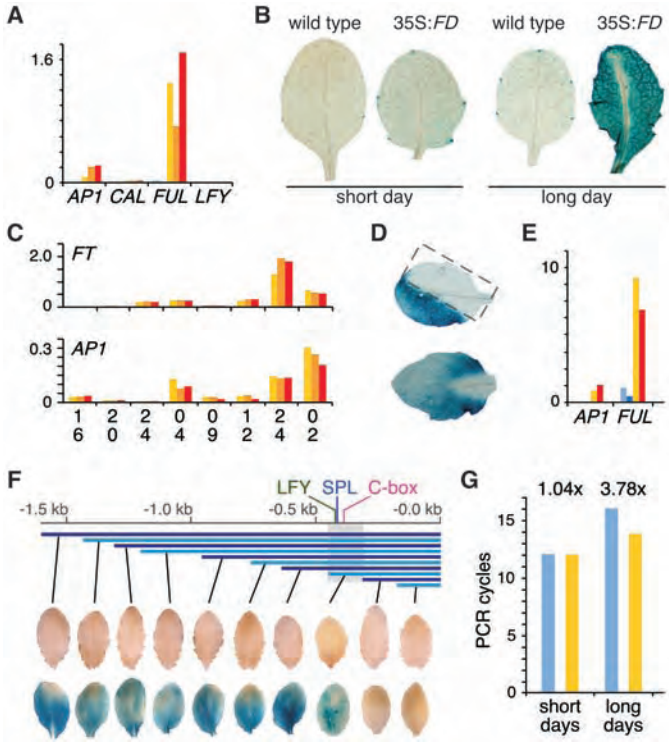


Fig. 4. Response of floral meristem-identity genes to FD and FT activity. (**A**) Expression of floral meristem-identity genes in leaves of 26-day-old, long-day-grown wild-type [left; two replicates, light and dark blue (mostly too small to be visible at this scale)] and 35S:*FD* plants (right; three replicates, yellow, orange, and red), as measured by real-time RT-PCR. *LFY* does not respond to 35S:*FD*. (**B**) Photoperiod-dependent activation of a *FUL:GUS* reporter in leaves of 35S:*FD*. (**C**) Activation of *FT* and *AP1* in leaves of 35S:*FD* plants in response to transfer to long days. Abscissa indicates chronological time, starting at 16:00 hours on the first day (end of short day). There appears to be a slight delay in *AP1* activation relative to *FT* activation. (**D**) *AP1:GUS* reporter activity is only induced in the part of a leaf from a 35S:*FD* plant that has been exposed to long days. Dashed box indicates covered portion of the leaf. An uncovered leaf is shown below. (**E**) Expression of *AP1* and *FUL* in 18-day-old 35S:*FT tfl1 ft* (left; two replicates, light and dark blue) and 35S:*FT-VP16 tfl1 ft* plants (right; two replicates, yellow and red). (**F**) Mapping of an *FD*-response element in the *AP1* promoter to a region that encompasses binding sites for LFY (*18*), SPL3 (*27*), and a C-box, a motif that can be bound by bZIP proteins (*28*). (**G**) Immunoprecipitation of chromatin from 35S:*FD* plants with antibodies to FT, followed by real-time PCR amplification of control *HSF1* promoter (blue) or *AP1* promoter (yellow) sequences (*11*). PCR cycle difference between precipitate and supernatant is plotted; a lower number of cycles indicates relative enrichment in the immunoprecipitate compared with the supernatant. Numbers on top indicate fold enrichment of *AP1* compared with *HSF1* sequences, by assuming PCR efficiency of 1.8-fold amplification per cycle. Results are averages of at least three replicates. Tissue was harvested at 23:15 hours on the second day after transfer from short to long days.

Table 1. Flowering times of mutant and transgenic plants. The 95% confidence interval is 2 × the standard error of the mean.

Genotype	Leaves	95% confidence interval	Range	*n*
		Experiment 1 (long days)		
Wild type	16.3	±0.7	15–18	12
*fd-2**	23.5	±0.6	18–28	54
		Experiment 2 (long days)		
Wild type	15.7	±0.8	10–19	23
35S:*FT*	5.3	±0.2	4–8	75
fd-2	22.8	±1.5	20–28	10
fd-2 35S:*FT*	10.7	±0.7	8–13	15
		Experiment 3 (long days)		
Wild type	15.6	±1.0	13–18	14
35S:*FD*	10.0	±0.4	8–12	24
		Experiment 4 (short days)		
Wild type	50.5	±0.4	48–53	30
35S:*FD*	45.6	±0.8	43–50	21
		Experiment 5 (long days)		
Wild type	13.6	±0.5	12–15	12
FD:FT (T1 plants)	5.3	±0.3	5–6	10

*At4g35900 insertion allele SALK_013288.

Two additional experiments demonstrated that ectopic activation of floral markers requires the combined presence of FT and FD. First, we manipulated the temporal expression of *FT* in 35S:*FD* plants by transferring them from short to long days. In this experimental set-up, activation of *AP1* closely follows that of *FT* (Fig. 4C), indicating that *AP1* is an early target of combined *FT* and *FD* activity. Second, we manipulated the spatial pattern of *FT* expression in 35S:*FD* plants containing an *AP1:GUS* reporter by partially covering leaves with aluminum foil. *FT* itself is cell-autonomously activated in response to light-dependent activation of the upstream regulator CO (*1*, *2*, *16*, *17*). Upon transfer of 35S:*FD AP1:GUS* plants from short to long days, which induces *FT* expression in leaves, *AP1:GUS* is strongly activated in the light-exposed but not the shaded part of a leaf (Fig. 4D).

How, then, does FT control FD activity? FT might affect posttranslational modification of FD, which in turn may alter its DNA binding capacity or transcriptional activity. Alternatively, FT itself might provide transcriptional activation potential. To test the latter scenario, we added the potent VP16 activation domain to FT and introduced a 35S:*FT-VP16* construct into *ft tfl1* double mutants. These plants flower even earlier than 35S:*FT ft tfl1*, and have a phenotype similar to that of 35S:*FT* 35S:*LFY* (*5*, *6*), with a single terminal flower between the first two leaves (Fig. 2E). *AP1* and *FUL* are expressed at much higher levels in 35S:*FT-VP16* than in 35S:*FT* plants, confirming that the VP16 domain alters FT activity (Fig. 4E).

That adding a transcriptional activation domain to FT changes its activity supports a model in which FT acts in the nucleus as part of a transcriptional complex with FD. To generate additional support for this scenario, we mapped the FD response element in the *AP1* promoter to a 130–base pair region (Fig. 4F), which also encompasses the previously identified binding site of LFY, another *AP1* activator (*18*). The LFY binding site is not required for ectopic activation of the *AP1* reporter, confirming an LFY-independent effect of FD on the floral transcriptional program. To test whether an FT/FD complex directly interacts with the FD response element, we performed chromatin immunoprecipitation, using 35S:*FD* plants and FT antibodies (*11*). *AP1* promoter sequences are enriched in long days, in which *FT* expression is induced, but not in short days (Fig. 4G). An enrichment of *AP1* sequences is not seen with leaf material of wild-type plants. Thus, FT protein is recruited to the FD-response element in the *AP1* promoter in an FD-dependent manner.

Roles of *FT* and *FD* in mediating floral inductive signals. It has long been known that day length is detected in leaves, which in

turn release a systemic signal dubbed "florigen" (*19*, *20*). Our data suggest a model in which FT directly regulates the activity of FD at the shoot apex by formation of a transcriptional complex, even though induction of *FT* RNA expression in response to photoperiod occurs predominantly in leaves. Consistent with *FT* being the primary target of CO in leaves, *CO* overexpression causes activation of the *FT* promoter in the phloem of leaves (*2*, *21*). The most parsimonious hypothesis to integrate these findings is that induction of *FT* RNA in the leaf phloem leads to the presence of FT protein at the shoot apex. Conventional means are insufficient to detect FT protein at the shoot apex, nor can *FT* RNA be detected at the shoot apex on microarrays (Fig. 1C). More sensitive real-time reverse transcription polymerase chain reaction (RT-PCR) analyses indicate a merely threefold increase of *FT* RNA at the shoot apex of plants grown in the long-day compared with the short-day condition; this contrasts with the approximately 70-fold difference in leaves (Fig. 1D). In addition, transgenic experiments show that *FT* activity at the shoot apex needs to be tightly controlled in order to prevent precocious flowering: When *FT* is expressed directly in the *FD* domain with the use of *FD* regulatory sequences, early flowering ensues, demonstrating exquisite sensitivity of the shoot apex to local *FT* levels (Table 1).

Our data do not necessarily imply that *FT* RNA or FT protein themselves are the long-distance signal. They are similarly consistent with a relay mechanism in which FT induces a mobile signal in leaves, which in turn leads to *FT* activation in secondary tissues near or at the shoot apex. Such an intermediate step may be required to translate the circadian activation of *FT* in leaves into a more constant and finely tuned signal at the shoot apex. Secondary activation of *FT* expression at the shoot apex may also explain why this tissue has an important function in perceiving the effects of vernalization, the transient exposure to winterlike temperatures that promotes flowering in many plants (*22*, *23*).

Finally, FT and FD are unlikely to be the only factors that mediate the effects of upstream regulators such as CO on flowering. Rather, we envision an intricate system of flowering promoting and repressing activities that act at the shoot apex to provide a robust flowering response. This system is likely to include the multitude of MADS box genes affecting flowering, such as *SOC1* and others, as well as the targets of the miR156 and miR172 microRNAs (*23*–*25*). How the FD/FT module interacts with these other factors and their role in transmitting the mobile flower-inducing signal are important avenues for further research.

References and Notes

1. F. Valverde *et al.*, *Science* **303**, 1003 (2004).
2. H. An *et al.*, *Development* **131**, 3615 (2004).
3. B. G. Ayre, R. Turgeon, *Plant Physiol.* **135**, 2271 (2004).
4. A. Samach *et al.*, *Science* **288**, 1613 (2000).
5. I. Kardailsky *et al.*, *Science* **286**, 1962 (1999).
6. Y. Kobayashi, H. Kaya, K. Goto, M. Iwabuchi, T. Araki, *Science* **286**, 1960 (1999).
7. M. J. Banfield, R. L. Brady, *J. Mol. Biol.* **297**, 1159 (2000).
8. W. J. Lemon, S. Liyanarachchi, M. You, *Genome Biol.* **4**, R67 (2003).
9. M. Schmid *et al.*, *Development* **130**, 6001 (2003).
10. L. Pnueli *et al.*, *Plant Cell* **13**, 2687 (2001).
11. Materials and methods are available as supporting material on *Science* Online.
12. M. Koornneef, C. J. Hanhart, J. H. van der Veen, *Mol. Gen. Genet.* **229**, 57 (1991).
13. M. Koornneef, C. Alonso-Blanco, H. Blankestijn-de Vries, C. J. Hanhart, A. J. M. Peeters, *Genetics* **148**, 885 (1998).
14. L. Ruiz-García *et al.*, *Plant Cell* **9**, 1921 (1997).
15. M. A. Mandel, M. F. Yanofsky, *Nature* **377**, 522 (1995).
16. P. Suárez-López *et al.*, *Nature* **410**, 1116 (2001).
17. M. J. Yanovsky, S. A. Kay, *Nature* **419**, 308 (2002).
18. F. Parcy, O. Nilsson, M. A. Busch, I. Lee, D. Weigel, *Nature* **395**, 561 (1998).
19. J. E. Knott, *Proc. Am. Soc. Hort. Sci.* **31**, 152 (1934).
20. J. A. D. Zeevaart, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **27**, 321 (1976).
21. S. Takada, K. Goto, *Plant Cell* **15**, 2856 (2003).
22. Y. He, R. M. Amasino, *Trends Plant Sci.* **10**, 30 (2005).
23. I. R. Henderson, C. Dean, *Development* **131**, 3829 (2004).
24. M. J. Aukerman, H. Sakai, *Plant Cell* **15**, 2730 (2003).
25. R. Schwab *et al.*, *Dev. Cell* **8**, 517 (2005).
26. M. Schmid *et al.*, *Nat. Genet.* **37**, 501 (2005).
27. G. H. Cardon, S. Hohmann, K. Nettesheim, H. Saedler, P. Huijser, *Plant J.* **12**, 367 (1997).
28. T. Izawa, R. Foster, N. H. Chua, *J. Mol. Biol.* **230**, 1131 (1993).

29. We thank T. Araki for discussing results before publication; the NSF-funded *Arabidopsis* Biological Resource Center and Nottingham *Arabidopsis* Stock Centre for supplying seeds of SALK insertion lines produced by J. Ecker and colleagues; N. Warthmann for proposing the leaf expression assays; and K. Bomblies, C. Dean, A. Maizel, J. Mylne, and R. Schwab for discussions and critical reading of the manuscript. Supported by fellowships from the Wellcome Foundation (P.W.), Korea Science and Engineering Foundation and European Molecular Biology Organisation (M.C.K.), a Deutsche Forschungsgemeinschaft *Arabidopsis* Functional Genomics Network grant (P.W.), a Human Frontier Science Program Organisation (HFSPO) Career Development Award (J.U.L.), an HFSPO research grant and the Bundesministerium für Bildung und Forschung Genomanalyse im biologischen System Pflanze program (D.W.), and the Max Planck Society. D.W. is a Director of the Max Planck Institute.

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5737/1056/DC1
Materials and Methods
Figs. S1 to S4
References

3 May 2005; accepted 7 July 2005
10.1126/science.1114358

REPORTS

Origin of Brittle Cleavage in Iridium

Marc J. Cawkwell,[1]* Duc Nguyen-Manh,[2] Christopher Woodward,[3,4] David G. Pettifor,[5] Vaclav Vitek[1]

Iridium is unique among the face-centered cubic metals in that it undergoes brittle cleavage after a period of plastic deformation under tensile stress. Atomistic simulation using a quantum-mechanically derived bond-order potential shows that in iridium, two core structures for the screw dislocation are possible: a glissile planar core and a metastable nonplanar core. Transformation between the two core structures is athermal and leads to exceptionally high rates of cross slip during plastic deformation. Associated with this athermal cross slip is an exponential increase in the dislocation density and strong work hardening from which brittle cleavage is a natural consequence.

The plastic deformation of crystalline materials is mediated by the motion of line defects called dislocations (*1*). Dislocations glide along crystallographic planes and displace the two parts of the crystal by a fixed lattice vector known as the Burgers vector. If metastable stacking faults can form, dislocations may reduce their energy by dissociating into partials with Burgers vectors smaller than the lattice vector, connected by the stacking fault. The equilibrium width of splitting is determined by a balance of elastic repulsion between the partials and attraction arising from the energy of the stacking fault. The mobility of individual dislocations is to a large extent governed by the shape of their cores. Dislocations with cores that are spread only on the slip plane are hereafter referred to as planar and are usually glissile; i.e., they move easily at very low applied stresses even at low temperatures. Dislocations with cores that are spread onto two or more nonparallel planes are hereafter referred to as nonplanar and tend to be sessile; i.e., in order to glide, they require high stresses that are often strongly dependent on temperature

(*2*). During plastic deformation, dislocations are continuously generated and multiplied by means of the operation of Frank-Read sources (*3*).

Dislocations in metals with the face-centered cubic (fcc) crystal structure commonly dissociate into Shockley partials and adopt glissile planar cores. The efficient multiplication of dislocations in fcc metals in conjunction with their high mobility ensures that any stress concentrators can be quickly relaxed. As a rule, such plasticity leads to failure by ductile tearing rather than brittle cleavage even at very low temperatures.

The refractory fcc metal iridium (melting point 2443°C) is a notable exception because it deforms plastically in a manner crystallographically consistent with other fcc metals but then undergoes brittle transgranular cleavage at temperatures up to 500°C under tension (*4*–*6*). This behavior is well established as being an intrinsic failure mode (*6*–*9*). The usual octahedral slip system is responsible for plastic deformation (*10*) and dislocations dissociate into Shockley partials (*11*). The feature that sets iridium apart from other fcc metals is its exceptionally high rate of work hardening during neck-free plastic deformation and the accumulation of dislocation densities so high that they can only be compared with other fcc metals after severe radiation damage (*6*). These features suggest that dislocation multiplication takes place in iridium at a rate far beyond that possible in other fcc metals.

We present a model for the very high rate of dislocation multiplication in iridium that leads ultimately to brittle cleavage. We demonstrate through atomistic simulations that (i) the screw dislocation in iridium can adopt a metastable nonplanar core in addition to a planar core that corresponds to dissociation into Shockley partials. (ii) Transformation from the planar core to the nonplanar core occurs under applied stresses that reduce the width of splitting between partials. It neither requires thermal activation nor full constriction of the partials. (iii) Once the nonplanar core is formed, it transforms back into Shockley partials on the cross-slip plane, again without the need for thermal activation. These three steps, all unique to iridium because of the strong angular character of its interatomic bonding, describe a mechanism through which cross slip may occur at a very high rate and without thermal activation. This athermal cross-slip mechanism leads to a precipitous increase in the number density of Frank-Read sources and to an exponentially increasing dislocation density. On reaching a certain level, high dislocation densities naturally lead to brittle cleavage because stress concentrations cannot be relaxed through further dislocation glide.

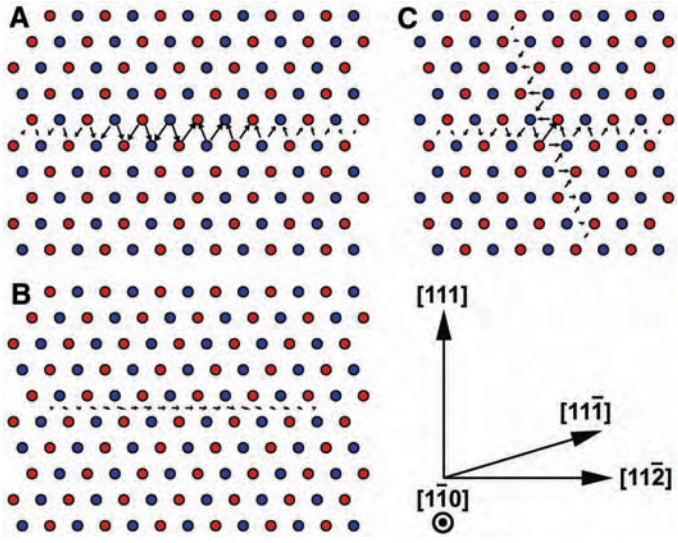


Fig. 1. Differential displacement maps of screw dislocation cores in fcc iridium. The circles depict atoms in the projection onto the plane perpendicular to the dislocation line, (1$\bar{1}$0), and the colors distinguish two planes of atoms within one period. The arrows represent relative displacement between two neighboring atoms. The length of the arrows is proportional to the magnitude of the displacement normalized by $a/2\sqrt{2}$, where a is the fcc lattice parameter. (**A**) Planar core with arrows representing displacements parallel to the dislocation line. (**B**) Planar core with arrows representing displacements perpendicular to the dislocation line. The dislocation is dissociated into Shockley partials and the region of arrows with constant length in (B) is the intrinsic stacking fault and its extent shows that the width of splitting is about 8 Å. (**C**) Nonplanar configuration with the dislocation core spreading into the intersecting (111) and (11$\bar{1}$) planes. Arrows represent displacements parallel to the dislocation line.

The development of an interatomic potential for fcc iridium requires that the angular character of bonding arising from the valence *d* electrons is captured accurately. We adopted the bond-order potential formalism, a real-space, $O(N)$, orthogonal tight-binding scheme (*12*–*14*), which has been shown to be eminently suitable for the atomistic simulation of extended defects in transition metals (*15*, *16*). Extensive testing of the bond-order potential developed for fcc iridium showed that it is an accurate and transferable description of bonding (*17*).

High-resolution transmission electron microscopy studies of screw dislocations in iridium showed that they dissociate into Shockley partials with a width of splitting of only 8 Å, corresponding to an intrinsic stacking fault energy of 420 mJ/m^2 (*11*). The bond-order potential yields an intrinsic stacking fault energy of 408 mJ/m^2. Our atomistic simulations of the screw dislocation in iridium were performed at zero temperature. Atomistic relaxation using the bond-order potential for iridium in conjunction with flexible lattice Green's function boundary conditions (*14*, *18*) found two core structures. The first is planar and corresponds to dissociation into Shockley partials on the (111) plane. It is presented as a differential displacement map in Fig. 1, A and B. The second structure (Fig. 1C) is nonplanar with the core spread into two intersecting {111} planes in the [1$\bar{1}$0] zone. Such a core structure has not been found in atomistic simulations of any other fcc metal. However, the ordinary screw dislocation in TiAl (tetragonal fcc-based structure) exhibits only this core structure because of the very high stacking fault energy (*19*).

In Fig. 2, we plot the energies of the dislocations with the planar and nonplanar cores calculated atomistically in concentric cylinders of radius R about their elastic center versus $\ln(R/b)$ [as per elasticity theory (*1*)], where b is the magnitude of the total Burgers vector. This shows that the dislocation with the nonplanar core is metastable and is higher in energy by 0.33 eV/b than the dislocation that is dissociated into Shockley partials.

The Burgers vectors of Shockley partials possess both screw and edge components, i.e., components parallel and perpendicular to the dislocation line respectively. Only shear stresses parallel to the total Burgers vector of the dislocation can drive its glide (glide stresses). Stresses perpendicular to the total Burgers vector, referred to as Escaig stresses, change only the separation of the partials (*20*, *21*). The dislocation with the nonplanar core is not affected by Escaig stresses because in this core structure, the edge components are negligible. Atomistic simulation of the application of glide stresses to the dislocation with the planar core showed that the minimum stress at which it glides (the Peierls stress) is 2×10^{-4} C_{44}, where C_{44} is a shear modulus for iridium. The low value of the Peierls stress verified that this dislocation is highly glissile.

[1]Department of Materials Science and Engineering, University of Pennsylvania, 3231 Walnut Street, Philadelphia, PA 19104–6202, USA. [2]European Atomic Energy Community/UK Atomic Energy Authority (EURATOM/UKAEA) Fusion Association, Culham Science Centre, Abingdon, OX14 3DB, UK. [3]Air Force Research Laboratory, Wright-Patterson Air Force Base, OH 45433–7817, USA. [4]Department of Materials Science and Engineering, Northwestern University, Evanston, IL 60208–3108, USA. [5]Department of Materials, University of Oxford, Parks Road, Oxford, OX1 3PH, UK.

*To whom correspondence should be addressed. E-mail: cawkwell@seas.upenn.edu

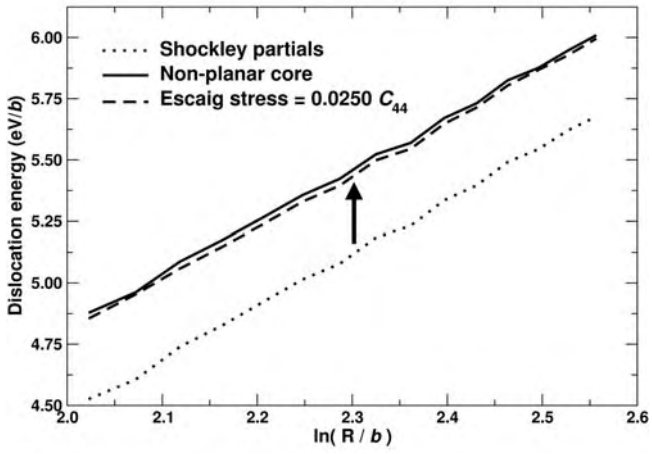


Fig. 2. Energy of screw dislocation with planar and nonplanar cores. The dashed line shows the energy of the dislocation dissociated into Shockley partials constricted by an Escaig stress of 0.0250 C_{44}.

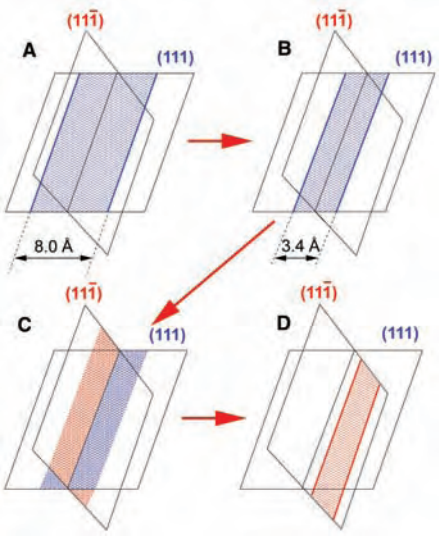


Fig. 3. Schematic diagram of athermal cross slip in iridium. (**A** and **B**) The splitting between Shockley partials is constricted from 8.0 to about 3.4 Å where the dislocations with the planar and nonplanar cores become energetically equivalent. (**C**) Long segments of the nonplanar core spontaneously form along the screw dislocation line. (**D**) The nonplanar core transforms under the effect of glide and Escaig stress into Shockley partials on the cross-slip plane.

We performed atomistic simulations of the change in energy of the dislocation with the planar core when the partials are constricted by the application of an Escaig stress. We found that at an Escaig stress of 0.0250 C_{44} the dislocations with the planar and nonplanar cores are energetically equivalent (Fig. 2). Elasticity calculations (*14*) estimated that the splitting between partials is reduced to 3.41 Å at this stress (figs. S2).

The calculated value of the Peierls stress of the dislocation with the planar core allowed us to estimate the likelihood of providing sufficient stress concentration to form the nonplanar core configuration. Considering a dislocation pile-up where the glide stress acting on each dislocation is equal to the Peierls stress, we estimated with the use of the standard elasticity treatment (*1*) that a maximum of 125 dislocations are required for the dislocation at the head of the pile-up to be sufficiently constricted. Because of the hardening effect of other crystal defects, the stress at which dislocation glide begins will be far higher than the Peierls stress, thus much smaller numbers of dislocations are needed in a pile-up for the transformation to the nonplanar core to occur.

Atomistic simulation of the application of glide stresses to the dislocation with the nonplanar core showed that it transforms into the dislocation with the planar core on the maximum resolved shear stress plane at a stress in excess of 8×10^{-5} C_{44}. This indicates that this core does not render the dislocation sessile, in contrast with other materials, in which dislocations with nonplanar cores are sessile and lead to high-yield stresses (*22*). However, we found that if this glide stress is applied together with an Escaig stress in excess of 5×10^{-5} C_{44} then the transformation always takes place onto the cross-slip plane (the intersecting {111} plane onto which the nonplanar core is spread). The Escaig stress affects the transformation by means of the edge components of the Burgers vectors of the Shockley partials as they evolve (*20*). If the partials are more constricted on the maximum resolved shear stress plane than on the cross-slip plane, it is energetically favorable for the nonplanar core to transform into the planar core on the cross-slip plane. However, to form the nonplanar core initially, substantial constriction of partials in the maximum resolved shear stress plane is necessary. We therefore conclude that once this core is formed, it immediately transforms onto the cross-slip plane (Fig. 3).

The small magnitude of the glide stress needed to transform the nonplanar core to the planar one suggests that the former is associated with a shallow energy minimum. This is a good indication that the activation barrier for the forward transformation from the constricted partials to the nonplanar core is small and thus occurs readily.

Our simulations show that in iridium, dislocation core structure transformations lead to cross slip without the need for thermal activation or constricting the partials fully in the primary plane. The consequences of such athermal cross slip, illustrated schematically in Fig. 3, for mechanical properties are considerable. This mechanism allows cross slip to occur over long segments of the screw dislocation line with great frequency, resulting in a high number density of Frank-Read sources generated by double cross slip (*23*). These sources are responsible for further generation of glissile dislocations that in turn cross slip to form new sources. Owing to this feedback mechanism, the rate of change of dislocation density, ρ, with plastic strain, ε, is at least linearly proportional to the dislocation density; i.e., $d\rho/d\varepsilon = C\rho$ and thus $\rho = \rho_0 \exp(C\varepsilon)$, where ρ_0 is an initial dislocation density and C a constant. Plastic strain can be written $\varepsilon = \rho b\bar{x}$ where $\bar{x}$ is the dislocation mean free path (*24*). It is evident that a rapidly increasing dislocation density during plastic deformation leads to small values of $\bar{x}$. If $\bar{x} \approx b$ then stress concentrations cannot be relaxed by dislocation glide and any cracks that pre-exist or form will propagate rather than blunt. Assuming that the strain at cleavage is around 10%, we estimate that the dislocation density is in excess of 10^{18} m^{-2}. This density is consistent with that found in Panfilov *et al*. (*6*). Our model for the origin of transgranular cleavage in the refractory fcc metal iridium is further reinforced by the experimental observation that thin iridium foils fail by plastic tearing (*25*). Thin foils are unable to accumulate such high dislocation densities as in the bulk because dislocations generated by means of the mechanism we describe escape through free surfaces.

The mechanism of athermal cross slip proposed here is likely to be unique to iridium because of its mixed metallic and covalent interatomic bonding, which is either absent or much weaker in other fcc metals. The nonplanar configuration for the screw dislocation core is unstable in other fcc metals but metastable in iridium because of the angular character of bonding arising from the valence *d* band. Similarly, the very high intrinsic stacking fault energy in iridium, which leads to a narrow splitting between partials and promotes the transformation from the planar core into the nonplanar one, is also directly related to the strong angular bonding.

References and Notes

1. J. P. Hirth, J. Lothe, *Theory of Dislocations* (Wiley-Interscience, New York, ed. 2, 1982).
2. V. Vitek, *Prog. Mater. Sci.* **36**, 1 (1992).
3. W. T. Read, *Dislocations in Crystals* (McGraw-Hill, New York, 1953).
4. P. Haasen, H. Hieber, B. L. Mordike, *Z. Metallkd.* **56**, 832 (1965).
5. C. A. Brookes, J. H. Greenwood, J. L. Routbort, *J. Inst. Metals* **98**, 27 (1970).
6. P. Panfilov, A. Yermakov, V. Dmitriev, N. Timofeev, *Platinum Metals Rev.* **35**, 196 (1991).

7. S. S. Hecker, D. L. Rohr, D. F. Stein, *Metall. Trans.* **9**, 481 (1978).
8. C. N. Reid, J. L. Routbort, *Metall. Trans.* **3**, 2257 (1972).
9. P. Panfilov, V. Novgorodov, A. Yermakov, *J. Mat. Sci. Lett.* **13**, 137 (1994).
10. R. Adamesku, S. Grebenkin, A. Yermakov, P. Panfilov, *J. Mat. Sci. Lett.* **13**, 865 (1994).
11. T. J. Balk, K. J. Hemker, *Philos. Mag. A* **81**, 1507 (2001).
12. D. G. Pettifor, *Phys. Rev. Lett.* **63**, 2480 (1989).
13. A. P. Horsfield, A. M. Bratkovsky, M. Fearn, D. G. Pettifor, M. Aoki, *Phys. Rev. B* **53**, 12694 (1996).
14. Materials and methods are available as supporting material on *Science* Online.
15. A. Girshick, A. M. Bratkovsky, D. G. Pettifor, V. Vitek, *Philos. Mag. A* **77**, 981 (1998).
16. M. Mrovec, D. Nguyen-Manh, D. G. Pettifor, V. Vitek, *Phys. Rev. B* **69**, 094115 (2004).
17. M. J. Cawkwell, D. Nguyen-Manh, D. G. Pettifor, V. Vitek, unpublished.
18. S. Rao, C. Hernandez, J. P. Simmons, T. A. Parthasarathy, C. Woodward, *Philos. Mag. A* **77**, 231 (1998).
19. C. Woodward, S. I. Rao, *Philos. Mag.* **84**, 401 (2004).
20. B. Escaig, *J. Phys. France* **29**, 225 (1968).
21. G. Lu, V. V. Bulatov, N. Kioussis, *Phys. Rev. B* **66**, 144103 (2002).
22. M. S. Duesbery, in *Dislocations in Solids*, F. R. N. Nabarro, Ed. (Elsevier, Amsterdam, 1989), vol. 8, pp. 67–173.
23. J. S. Koehler, *Phys. Rev.* **86**, 52 (1952).
24. D. Hull, D. J. Bacon, *Introduction to Dislocations* (Pergamon, Oxford, ed. 3, 1984).
25. P. Panfilov, in *Iridium*, E. K. Ohriner, R. D. Lanam, P. Panfilov, H. Harada, Eds. (TMS, Warrendale, PA, 2000), pp. 27–39.
26. This work was supported by the U.S. Department of Energy BES grant no. DE-PG02-98ER45702 (M.J.C. and V.V.), the Air Force Research Laboratory, the Air Force Office of Scientific Research under contract F33615-01-C-5214 (C.W.) and the UK Engineering and Physical Sciences Research Council and EURATOM (D.N.M.). We are grateful to M. Khantha, C. J. McMahon Jr., A. P. Sutton, and M. Mrovec for many fruitful discussions.

Supporting Online Material
www.sciencemag.org/cgi/content/full/309/5737/1059/DC1
Materials and Methods
Figs. S1 and S2
References

11 May 2005; accepted 23 June 2005
10.1126/science.1114704

Photochemical Mass-Independent Sulfur Isotopes in Achondritic Meteorites

Vinai K. Rai,* Teresa L. Jackson, Mark H. Thiemens

Sulfides from four achondrite meteorite groups are enriched in ^{33}S (up to 0.040 per mil) as compared with primitive chondrites and terrestrial standards. Stellar nucleosynthesis and cosmic ray spallation are ruled out as causes of the anomaly, but photochemical reactions in the early solar nebula could produce the isotopic composition. The large ^{33}S excess present in oldhamite from the Norton County aubrite (0.161 per mil) suggests that refractory sulfide minerals condensed from a nebular gas with an enhanced carbon-oxygen ratio, but otherwise solar composition is the carrier. The presence of a mass-independent sulfur effect in meteorites argues for a similar process that could account for oxygen isotopic anomalies observed in refractory inclusions in primitive chondrites.

In 1973, an oxygen isotopic anomaly was reported in the refractory calcium-aluminum-rich inclusions from the primitive chondrite Allende. It was suggested that this isotopic signature arose from the admixture of ^{16}O-rich exotic dust with isotopically normal oxygen in the solar nebula (*1*). Despite searches for similar anomalies (*2*) in other elements, as expected for a nucleosynthetic source, none has been observed (*3*), implying that the observed oxygen anomaly in meteorites is produced by chemical rather than nuclear processes, perhaps by isotopic self-shielding of CO in the solar nebula (*4*) or chemically produced mass-independent (MI) isotopic anomalies arising from quantum-based symmetry effects (*5*–*8*).

Other than oxygen, sulfur is the only multi-isotopic system that can potentially record MI processes in the early solar nebula, because it possesses both a multi-isotopic composition as well as a relevant gas phase chemical speciation. From laboratory experiments, it has been shown that ultraviolet (UV) photolysis of a variety of sulfur-containing compounds of potential importance in either the solar nebula or early planetary atmospheres is capable of producing MI isotopic compositions (*9*). Despite observation of MI sulfur in a variety of natural (*10*, *11*) and laboratory photolysis experiments (*9*), sulfur in primitive chondrites showed a narrow range of mass-dependent isotopic variation (*12*, *13*). In this paper, we report sulfur isotopes in achondrites that could have importance in delineation of physicochemical conditions in the early solar nebula and the origin of sulfur and potentially oxygen isotopic compositions.

We measured the sulfur isotopic composition of sulfide phases (*14*) in four groups of achondrites: howardite-eucrite-diogenite (HEDs), acapulcoite-lodranites, aubrites, ureilites (Fig. 1), and two ungrouped meteorites (Bencubbin and Weatherford), which carry heavy nitrogen. The $\delta^{34}S$ of achondrites fall within the ranges of previously analyzed carbonaceous chondrites, ordinary chondrites, and enstatite chondrites (*12*, *13*). Of the four classes, three of them have a significant enrichment of ^{33}S as indicated by positive $\Delta^{33}S$ values (*15*). The average value of $\Delta^{33}S$ for HEDs is 0.036 ± 0.016 per mil (‰) (standard error of six analyses); for acapulcoite-lodranite, 0.026 ± 0.008‰ (seven analyses); and for aubrites, 0.015 ± 0.006‰ (six analyses) (Table 1). After addition of two more ureilite samples to the data set of (*16*), the mean of $\Delta^{33}S$ is 0.040 ± 0.006‰ (standard error of 23 measurements). The sulfur phases in HEDs, acapulcoite-lodranites, and ureilites have clear, resolvable excesses of ^{33}S over analytical uncertainty as compared with measurements of a troilite from Canyon Diablo (CDT), which fall close to zero (average value of $\Delta^{33}S$ is 0.001 ± 0.002‰, 11 measurements) (*22*). The oldhamite separate from the Norton County aubrite possesses a significantly greater ^{33}S excess, with an average $\Delta^{33}S$ value of 0.161 ± 0.012‰; the mean value of aubrites $\Delta^{33}S$ (0.015 ± 0.006‰) is closer to the mass-dependent fractionation line compared with other achondrites, though is significantly resolved. The average $\Delta^{33}S$ of all achondrite data together (0.040 ± 0.006‰) is different from the average of sulfides from chondrites (0.002 ± 0.004‰) (Fig. 1), whereas the corresponding averages of $\Delta^{36}S$ are –0.013 ± 0.056‰ and 0.036 ± 0.028‰, the same within analytical uncertainty.

The enrichment of ^{33}S in achondrites is small compared with the observed oxygen isotopic anomaly of primitive chondrites. This may be because oxygen is carried in highly refractory and acid-resistant oxides, which may allow its production and preservation in nebular and parent body processes. Sulfur is carried by relatively volatile phases and therefore prone to isotopic homogenization. In addition, oxygen has multiple mechanisms by which MI isotopic compositions may be generated, both photochemical and thermal-chemical (*5*, *17*). The ^{33}S enrichment is roughly anticorrelated with the amount of sulfur, indicating that anomalous sulfur is contained by a relatively minor phase whereas most of the total sulfur is isotopically normal. This can be seen in the Norton County aubrite that has a normal bulk isotopic composition, whereas an oldhamite separate possesses a large excess of ^{33}S. Anomalous sulfur might also be present in primitive chondrites as well; however, the presence of large amounts

Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093–0356, USA.

*To whom correspondence should be addressed. E-mail: rai@chem.ucsd.edu

of normal sulfur and similar, if not identical, chemical behavior during extraction mask the more primitive anomalous component. A ^{33}S anomaly has been reported in an acid residue from Allende but a subsequent attempt to duplicate it failed (*18*), also suggesting that the carrier is a small, heterogeneous component.

Several processes could explain the origin of MI sulfur observed in achondrites: (i) cosmic ray spallation, (ii) stellar nucleosynthesis, and (iii) gas phase chemical reactions in the solar nebula, both chemical and photochemical. Spallation is ruled out because it produces a nearly constant ratio of $^{36}S/^{33}S \sim 6$ (or $\Delta^{36}S/\Delta^{33}S \sim 8$) in the metallic phase of iron meteorites (*19*). Furthermore, unlike iron meteorites, which have both higher concentrations of the target element (Fe in the case of S) as well as the requisite higher cosmic ray exposure ages (CREAs), the achondrites have insufficient iron and most of them have too low CREAs (e.g., acapulcoite-lodranites have CREAs < 10 Ma). Spallation on Ca as a target also produces more ^{36}S than ^{33}S enrichment. Production by thermal neutron capture on ^{32}S requires irradiation in a large parent body and high CREAs. The Norton County aubrite has both a large preatmospheric size as well as higher CREA, and thermal neutron capture of ^{32}S is possible in this meteorite. If this occurred, this would have produced uniform enrichments of ^{33}S in oldhamite as well as other sulfur phases. The lower $\Delta^{33}S$ in bulk Norton County as compared with that of oldhamite argues against this mode of origin.

Sulfur isotopes are synthesized in different stellar environments: ^{32}S and ^{34}S are made from hydrostatic and explosive oxygen burning, whereas ^{33}S is produced from explosive oxygen and neon burning (*20*–*22*). On the other hand, ^{36}S is produced in the convective shell in C-burning massive stars under hydrostatic conditions before the supernova (SN) II explosion (*23*). ^{36}S is also the least abundant sulfur isotope, and therefore any inhomogeneity due to incomplete mixing of different nucleosynthetic components of sulfur would be more apparent in $\delta^{36}S$ variations. Within the analytical uncertainty, the $^{36}S/^{32}S$ of all the meteorites is quite constant. Therefore, we rule out the possibility that the ^{33}S enrichment is of nucleosynthetic origin.

The only alternative remaining is that the MI sulfur in achondrites is produced by gas-phase MI chemistry or photochemistry in the solar nebula. The source of the fractionation derives from symmetry-dependent factors during chemical reactions (*5*). In the symmetry-driven model, one expects a similar effect in ^{36}S. The reaction CS + S might be considered, although it is unlikely because the presumed steady-state concentration of the reactants was likely low and the enrichments in ^{34}S and ^{36}S should track ^{33}S, which they do not. Within the error of the measurements, it is not possible to determine if such an isotopic correlation exists; however, the lack of a relevant chemical reaction places the largest constraint on sulfur.

An MI isotopic composition for sulfur and oxygen isotopes can be generated in the Earth and Martian atmospheres (*11*, *24*, *25*) and subsequently transferred to surface rocks. MI sulfur has also been observed in sulfonic acid extracts from the Murchison meteorite (*26*) and in photopolymerization of CS and CS_2 (*27*) and photolysis of SO_2 and H_2S in laboratory experiments (*9*, *11*). The HED, acapulcoite-lodranite, and ureilite data points plot close to the SO_2 photolysis fractionation pattern observed with a high-pressure Xe lamp (spectral region > 220 nm) (Fig. 2), but SO_2 is an unlikely species in early solar nebula. For

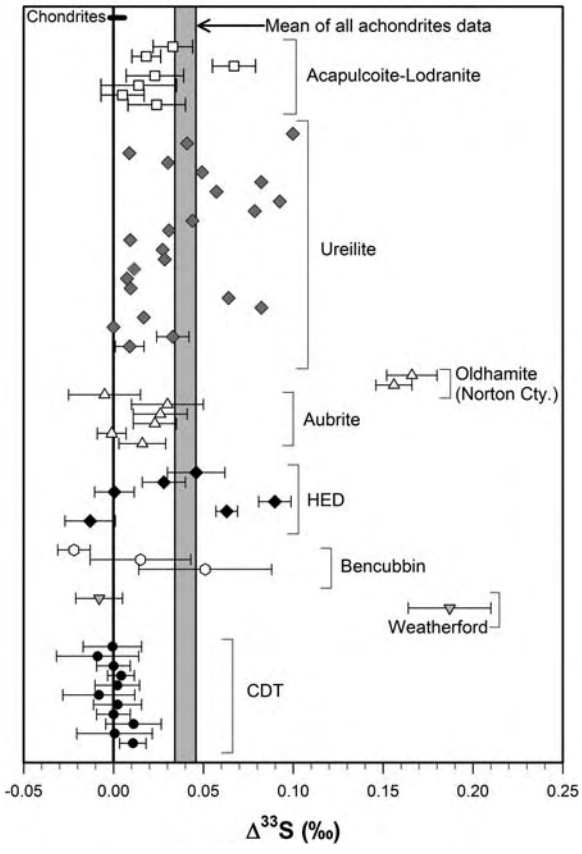


Fig. 1. The $\Delta^{33}S$ values from sulfide phases of individual samples of achondrites. Most of the data of ureilites (except two with error bars) are from (*16*). To demonstrate the precision of our measurements, we also show $\Delta^{33}S$ of CDT for reference. Most of the data points have positive $\Delta^{33}S$. The shaded box is the average of all the $\Delta^{33}S$ of sulfide phases from achondrites. Average $\Delta^{33}S$ of all the chondrites data are also shown at the top of the plot (*12*, *13*). Error bars indicate SE.

Table 1. Isotopic composition of acid-volatile sulfur from achondrites (this work) and chondrites (*12*, *13*). All the δ values are relative to CDT. Mean percentage weight of sulfur present as sulfide phases. The error reported here is 1 σ. Second numbers in parentheses are standard errors of the mean; *n* is number of analyses. Blank entries are not calculated.

	Mean wt%	Mean $\delta^{34}S$ (SEM)	$\delta^{34}S$ range (‰)	Mean $\Delta^{33}S(n)$ (‰) (SEM)	Mean $\Delta^{36}S(n)$ (‰) (SEM)
			Achondrites (this work)		
Acapulcoite-lodranite	0.95	0.015 ± 0.147 (0.056)	–0.144–0.237	0.026 ± 0.020(7) (0.008)	–0.002 ± 0.334(7) (0.126)
HED	0.10	0.373 ± 0.170 (0.069)	0.242–0.691	0.036 ± 0.039(6) (0.016)	–0.259 ± 0.622(6) (0.254)
Aubrite	0.41	0.319 ± 0.747 (0.305)	–0.577–1.710	0.015 ± 0.039(6) (0.006)	0.182 ± 0.106(6) (0.043)
Mean oldhamite (Norton County)	20	0.216 ± 0.012	0.063–0.119	0.161 ± 0.012	–0.020 ± 0.264
Ureilite	0.09	0.457 ± 0.293 (0.061)	–0.020–1.050	0.040 ± 0.031(23) (0.006)	–0.031 ± 0.338(23) (0.072)
			Chondrites		
Carbonaceous chondrites			–7.32–6.05	0.012 ± 0.027(15) (0.007)	–0.027±.148(8) (0.053)
Ordinary chondrites			–0.12–0.59	0.005 ± 0.028(14) (0.007)	0.278 ± 0.131(4) (0.066)
Enstatite chondrites			–0.39– –0.16	–0.007 ± 0.014(8) (0.005)	–0.049 ± 0.067(5) (0.030)

photolysis experiments with H_2S, a ^{33}S depletion rather than enrichment is observed in the products, suggesting that the residual H_2S is responsible for meteoritic observations. In all photolysis experiments except H_2S, MI effects in ^{33}S are accompanied by larger effects in ^{36}S. In H_2S photolysis, effects in ^{33}S are significantly larger than analytical uncertainty, whereas the effect on ^{36}S is normal, within measurement precision (*11*). At temperatures above 900 K, SiS replaces H_2S as primary S-bearing species in reducing condition in the nebula (*29*), but no photochemical fractionation data is available to assess the role of this in ^{33}S enrichment. For other gases that have been studied, such as H_2S, CS_2, and OCS, further laboratory experiments at extended wavelengths will be important in the future.

Oldhamite from the Norton County aubrite appears to be a potential carrier phase. Oldhamite is condensed from a gas more reduced than solar gases (C/O $\geq$ 0.95) but otherwise is of solar composition (*30*) and carries nearly all the rare earth elements in enstatite chondrites or aubrites (*31*). Such a high C/O ratio in the solar nebula can be achieved locally either by removal of H_2O or evaporation of carbon-rich materials of presolar origin (*32*, *33*). Oldhamite and other minerals condensed in a reducing environment have been identified in ureilites (*34*). Some of the sulfur phases in other meteorites might derive their MI sulfur from other refractory sulfide condensates formed in a nebular region with enhanced C/O ratio.

The presence of MI sulfur in four major achondrite groups that formed in distinct nebular regions either requires a widespread mechanism by which MI sulfur is produced or generated in restricted locales and subsequently transported to the nebular regions where these meteorite groups accreted. Considering the optical opacity of the accretionary disk, necessity of high-energy UV photons for production of MI sulfur, and high condensation temperature of oldhamite, a highly energetic environment near the protosun is a probable place. In the x-wind model (*35*, *36*) for protostellar evolution, precursor gases of sulfur-rich minerals experience UV light along the edges of the disk near the protosun and the x-region by high-energy photons from the embedded and revealed stages of the protosun (Fig. 3). Reduced conditions occur in these regions either by evaporation of presolar carbonaceous matter or by diffusive water vapor redistribution (*29*). In this region, the gas and dust condense and accrete to form fluffy grains, which are then moved into the x-region. Later on, these grains are carried by the x-wind to various nebular regions where they are incorporated into meteorite parent bodies. Although the observations are consistent with the x-wind model, a similar nebular scenario might account for the production of the sulfur anomalies, but the same parameters are required.

Because these phases are unstable under oxidizing conditions (solar C/O ratio), many of them undergo secondary reactions and lose their original MI signature. Other sulfide minerals also appear in the condensation sequence at lower temperatures (<700 K) from a gas of solar C/O ratio ($\leq$0.45) (*37*), which might not have experienced sufficient UV light exposure and hence do not have a MI signature. Mixing of these sulfides with refractory, and likely minor amounts of, sulfides produced in reducing conditions is the most likely reason why MI sulfur is not apparent in chondrites. Once they are inside the parent body of achondrites, oldhamite and accompanied sulfides behave like refractory minerals and survived preferentially during the parent body melting and recrystallization.

Sulfur is the only multi-isotopic system other than oxygen where MI fractionation has been observed in both natural environments and laboratory experiments. In a gas of solar composition, the first condensates in the early solar nebula are refractory oxides, whereas under reducing conditions with enhanced C/O ratios the refractory oxides are replaced by refractory sulfides in the condensation sequence, i.e., they are contemporary and might have experienced similar physicochemical conditions. The presence of MI compositions in these two refractory oxides and sulfides and the chemical origin of MI sulfur in meteorites suggest a similar origin for oxygen isotope fractionation in meteorites.

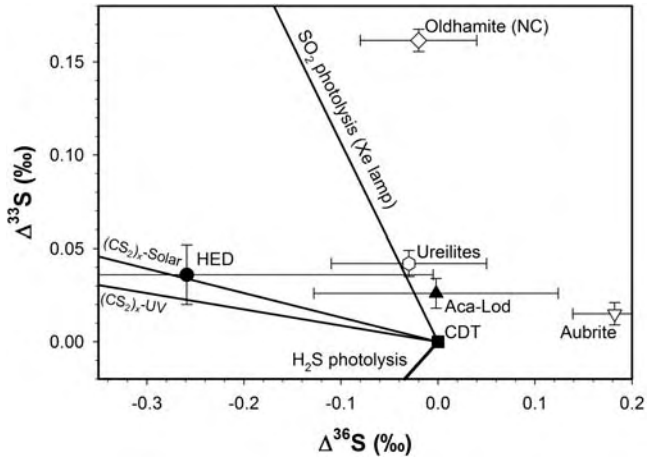


Fig. 2. Plot of mean $\Delta^{33}S$ and $\Delta^{36}S$ of various achondrite groups. Mass-dependent sulfur on this plot should plot close to (0,0). Various lines shown are obtained by mixing chondritic sulfur and MI sulfur obtained in laboratory experiments with the use of different UV lamps (*9*, *11*, *27*). Error bars indicate SE; NC, Norton County.

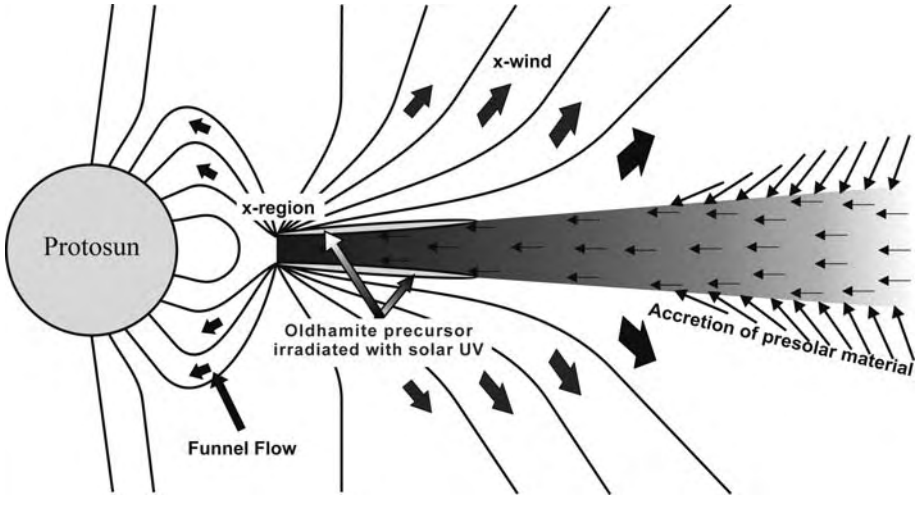


Fig. 3. Schematic drawing of x-wind model [reproduced from (*36*)]. The edges of disk near protosun, which can see direct UV light, are the regions where the refractory sulfide minerals were irradiated with solar UV radiation, which produced MI fractionation of sulfur in achondrites.

References and Notes

1. R. N. Clayton, L. Grossman, T. K. Mayeda, *Science* **182**, 485 (1973).
2. L. R. Nittler, C. M. O. Alexander, *Geochim. Cosmochim. Acta* **67**, 4961 (2003).
3. There have been numerous presolar oxide grains with anomalous oxygen identified, but their abundance is far too small and of the wrong isotopic composition to account for oxygen isotopic anomalies known to be present at the bulk meteorite level [(*2*) and references therein].
4. M. H. Thiemens, J. E. Heidenreich, *Science* **219**, 1073 (1983).
5. Y. Q. Gao, R. A. Marcus, *Science* **293**, 259 (2001).
6. M. H. Thiemens, *Science* **283**, 341 (1999).
7. M. H. Thiemens, J. Savarino, J. Farquhar, H. Bao, *Acc. Chem. Res.* **34**, 645 (2001).
8. M. H. Thiemens, in *Treaties on Geochemistry*, R. F. Keeling, Ed. (Elsevier, Oxford, 2003), vol. 4, pp. 159–175.

9. J. Farquhar, J. Savarino, S. Airieau, M. H. Thiemens, *J. Geophys. Res.* **106**, 32829 (2001).
10. J. Farquhar *et al.*, *Science* **298**, 2369 (2002).
11. J. Farquhar, J. Savarino, T. L. Jackson, M. H. Thiemens, *Nature* **404**, 50 (2000).
12. X. Gao, M. H. Thiemens, *Geochim. Cosmochim. Acta* **57**, 3159 (1993).
13. X. Gao, M. H. Thiemens, *Geochim. Cosmochim. Acta* **57**, 3171 (1993).
14. Sulfide extractions were done by using the procedure described in (*13*); also see online Materials and Methods for more details. The typical errors of $\delta^{33}S$, $\delta^{34}S$, and $\delta^{36}S$ measurements are 0.010, 0.010, and 0.200 (in ‰), respectively.
15. We calculated ^{33}S and ^{36}S enrichment with the following equation: $\Delta^{33}S = \delta^{33}S - 1000[(1 + \delta^{34}S/1000)^{0.515} - 1]$ and $\Delta^{36}S = \delta^{36}S - 1000[(1 + \delta^{34}S/1000)^{1.91} - 1]$.
16. J. Farquhar, T. L. Jackson, M. H. Thiemens, *Geochim. Cosmochim. Acta* **64**, 1819 (2000).
17. M. F. Miller *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10988 (2002).
18. C. E. Rees, H. G. Thode, *Geochim. Cosmochim. Acta* **41**, 1679 (1977).
19. X. Gao, M. H. Thiemens, *Geochim. Cosmochim. Acta* **55**, 2671 (1991).
20. Y. N. Chin, C. Henkel, J. B. Whiteoak, N. Langer, E. B. Churchwell, *Astron. Astrophys.* **305**, 960 (1996).
21. S. E. Woosley, A. Heger, T. A. Weaver, *Rev. Mod. Phys.* **74**, 1015 (2002).
22. R. Mauersberger, U. Ott, C. Henkel, J. Cernicharo, R. Gallino, *Astron. Astrophys.* **426**, 219 (2004).
23. S. E. Woosley, T. A. Weaver, *Astrophys. J. Suppl. Ser.* **101**, 181 (1995).
24. J. Farquhar, M. H. Thiemens, T. Jackson, *Science* **280**, 1580 (1998).
25. J. Farquhar, H. Bao, M. Thiemens, *Science* **289**, 756 (2000).
26. G. W. Cooper, M. H. Thiemens, T. L. Jackson, S. Chang, *Science* **277**, 1072 (1997).
27. J. J. Colman, X. Xu, M. H. Thiemens, W. C. Trogler, *Science* **273**, 774 (1996).
28. As a result of its small abundance, $\delta^{36}S$ is sensitive to contaminations at any stage from chemical extraction or gas chromatography or during measurements. The reported error on $\delta^{36}S$ is only measurement error; the actual uncertainty due to contamination might be higher.
29. M. A. Pasek *et al.*, *Icarus* **175**, 1 (2005).
30. J. W. Larimer, M. Bartholomay, *Geochim. Cosmochim. Acta* **43**, 1455 (1979).
31. K. Lodders, B. Fegley, *Earth Planet. Sci. Lett.* **117**, 125 (1993).
32. A. N. Krot, B. Fegley, K. Lodders, H. Palme, in *Protostars and Planets IV*, V. Mannings, A. P. Boss, S. S. Russell, Eds. (Univ. of Arizona, Tucson, AZ, 2000), pp. 1019–1054.
33. H. Nakano, A. Kouchi, S. Tachibana, A. Tsuchiyama, *Astrophys. J.* **592**, 1252 (2003).
34. P. Ramdohr, *Meteoritics* **7**, 565 (1972).
35. F. H. Shu, H. Shang, T. Lee, *Science* **271**, 1545 (1996).
36. F. H. Shu, H. Shang, A. E. Glassgold, T. Lee, *Science* **277**, 1475 (1997).
37. L. Grossman, *Geochim. Cosmochim. Acta* **36**, 597 (1972).
38. NASA Cosmochemistry program is gratefully acknowledged for support of this research.

Supporting Online Material
www.sciencemag.org/cgi/content/full/309/5737/1062/DC1
Material and Methods
Table S1
References and Notes

30 March 2005; accepted 12 July 2005
10.1126/science.1112954

Khipu Accounting in Ancient Peru

Gary Urton and Carrie J. Brezine

Khipu are knotted-string devices that were used for bureaucratic recording and communication in the Inka Empire. We recently undertook a computer analysis of 21 khipu from the Inka administrative center of Puruchuco, on the central coast of Peru. Results indicate that this khipu archive exemplifies the way in which census and tribute data were synthesized, manipulated, and transferred between different accounting levels in the Inka administrative system.

Tribute in the Inka state was levied in the form of a labor tax. Each "taxpayer" (state laborer) was required to work a specified number of days each year on state projects. Using data recorded in khipu (knotted-string devices used for bureaucratic recording and communication), Inka accountants assessed tribute levels and assigned tasks to different numbers of local workers. At the lowest, local level of the administrative hierarchy, tributaries were grouped into five accounting units of 10 members each. One member of each of these groups of 10 would have served as Chunka Kamayoq ("organizer of 10"). Five such groupings would make a unit of 50 tribute payers, under the authority of a Pichqa-Chunka Kuraka ("lord of 50"). Two groups of 50 would be combined into a unit of 100 tributaries led by a Pachaka Kuraka ("lord of 100") and so on up the hierarchy.

Near the top of the decimal administrative hierarchy were the heads of the approximately 80 provinces, the officials of which were called T'oqrikoq. Each provincial official was under the direction of the appropriate Lord of the Four Quarters; these four lords served directly under the Inka king in Cusco. The governor of each province was required to keep a copy of khipu accounts so that "no deception could be practiced by either the Indian tribute payers or the official collectors" (*1*).

Department of Anthropology, Harvard University, Cambridge, MA 02138, USA.

A primary question is how did information move between adjacent levels of this hierarchical administration? The instructions of higher-level officials for lower-level ones would have moved, via khipu, from the top of the hierarchy down. This information would be partitive in nature; for instance, assignments made to 1000 tribute payers would be broken down into two groups of 500, each of which would be decomposed into five groups of 100, and so on. In the reverse direction, local accountants would pass data regarding accomplished tasks upward through the hierarchy. In that direction, information at each level would represent the summation of accounts from the level immediately below. These accumulating data would eventually arrive in the hands of the Cusco accountants, where the highest level of accounting went on. Here we present an analysis of a set of khipu from Puruchuco that are linked hierarchically in such a relationship of summation and partition.

The archaeological site of Puruchuco is located on the south bank of the Rimac River, about 11.5 km northeast of the center of Lima, within the present-day district of Ate. Puruchuco is a roughly rectangular compound with high surrounding walls made of tapia (pounded adobe) construction. Around and in some cases abutted to the palace of Puruchuco were several smaller constructions. The cache of khipu was found under the floor of one of the smaller attached buildings. From its location, Mackey surmised that this building was the house of a khipu-keeper (khipukamayuq) who served the lord of the palace (*2*). Field notes from the day on which the khipu were discovered state that they were found inside a semi-ovoid urn covered by a small gourd. There were 21 khipu and several loose pendant strings (*3*).

What we term the Puruchuco "accounting hierarchy" pertains to 7 of the 21 khipu samples found together in the urn. Though not included in this analysis, several other khipu may provide supporting documentation to these seven. The seven khipu are related in a hierarchical arrangement of three interconnected levels, designated levels I, II, and III, as shown in Fig. 1. Two of the seven khipu (UR63 and UR73) were on level I, the base; three khipu were on the second level [UR64, UR68, and 9 (*4*)]; and two (UR67 and UR66) were on level III.

The two samples at the top of the hierarchy, UR66 and UR67, were rolled up together into a single bundle. These two khipu bear identical numerical values and string colors that seem to be a subtle transformation from one to the other.

There are two principal aspects of the Puruchuco accounting hierarchy. First, khipu on the same level match or closely match: They display identical or similar numerical sequences and color patterning. This, we argue, was the checks-and-balances aspect of the accounting hierarchy. Second, values on khipu sum upward and are subdivided downward: The numerical values of certain groupings of strings (to be defined below) on the two khipu on level I sum to values tied onto certain groupings of strings on the three khipu on level II, and the numerical values of certain groupings of strings on the three khipu on level II sum to the values on the two khipu on level III. Or, moving down the hierarchy, values on strings at higher levels are partitioned among groupings of strings on the next lowest level.

Through cord color and spacing, each of the seven khipu is organized into different

numbers of subunits. Khipu on level I decompose into six subunits; those on level II contain three subunits (plus what we call "introductory segments"); and the two khipu on level III have only one unit (plus introductory segments). Inside these subunits, the strings are further subdivided by a combination of spacing between strings and/or by the repetition of color patterning in groups of strings. The general color pattern is a four-string seriation or sequence of colors (such as dark brown, medium brown, light brown, and white) repeated multiple times (*5*, *6*). The numerical values of the cords vary in magnitude in accordance with the color, with the four strings of each color-seriated set generally increasing in size through the sequence.

An example of summation upward, between UR68 on level I and UR63 on level II, is given in Fig. 2. UR63 is organized by spacing and color seriation into six pendant string groupings, labeled a to f. The number of strings in each group is shown in brackets at the bottom of the columns. The six columns comprise (i) three sets of ($5 \times 4 =$) 20 strings organized into five groups of four color-seriated strings; (ii) two sets of ($3 \times 4 + 2 \times 3 =$) 18 color-seriated strings; and (iii) one set of ($3 \times 4 + 3 =$) 15 color-seriated strings. The meandering dotted lines at the tops and bottoms of the columns of UR63 in Fig. 2 show how this sample is to be reassembled into its proper linear arrangement. The numerical values of string groupings in UR63 sum to values recorded on the middle of the three subunits of UR68. The color-seriated strings of UR63 are aligned across the six segments, and these groupings are aligned with the similarly color-seriated grouping of ($5 \times 4 =$) 20 strings in the central subdivision (strings 34 to 53) of khipu UR68. Summing across the aligned strings of UR63 results in totals equal or close to those recorded on the depicted section of UR068. The values knotted into the cords of UR68 are reported on the right; any number between parentheses immediately to the left of these is the actual sum of values on the strings of UR63 at that position. The parenthetical numbers represent values that should have been recorded if the relationship between UR63 and UR68 was a matter of strict addition. The presence of several close, rather than exact, matches suggests that there was some degree of flexibility allowable in the accounting relationship between these two levels.

Continuing the summing upward, we next consider khipu UR68 (level II) and UR67 (level III). Their relationship is illustrated in Fig. 3. UR68 is disassembled into its three color-seriated subdivisions (labeled A to C), which are shown aligned with the similarly color-seriated string groupings of UR67. Figure 3 shows 20 strings in all subunits.

The summations between UR68 and UR67 are more exact than those between UR63 and UR68. Setting aside the broken string in UR67, the values diverge in only two instances, and in each case the discrepancies are small: 2904 instead of 2908 and 161 instead of 162. The variance present in the connection between levels I and II has been considerably reduced between levels II and III.

Pendants between dotted lines in Fig. 1 are implicated in the summation/partition relationship. The pendants on level III outside of the dotted lines, and those to the left of the dotted lines that protrude from the tops of the khipu on level II, form introductory segments. The dotted lines in Fig. 1 encompass all the pendants on level I khipu but only the middle subunit of level II khipu. That is, complete summation of level I khipu accounts for only a portion of the values recorded on khipu on level II. The other values on level II khipu are not accounted for by the currently known level I khipu UR63 and UR73. There may have been four additional level I khipu, with the information for these two additional subunits on level II. One pair would have summed to the leftmost subunits on level II, whereas the other would have produced sums recorded on

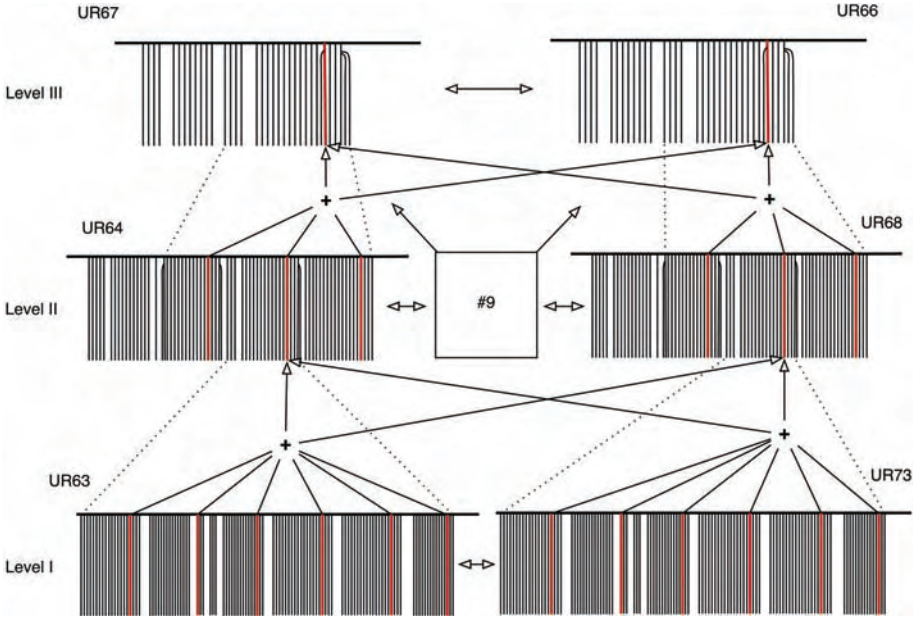


Fig. 1. The accounting hierarchy from the archive of Puruchuco.

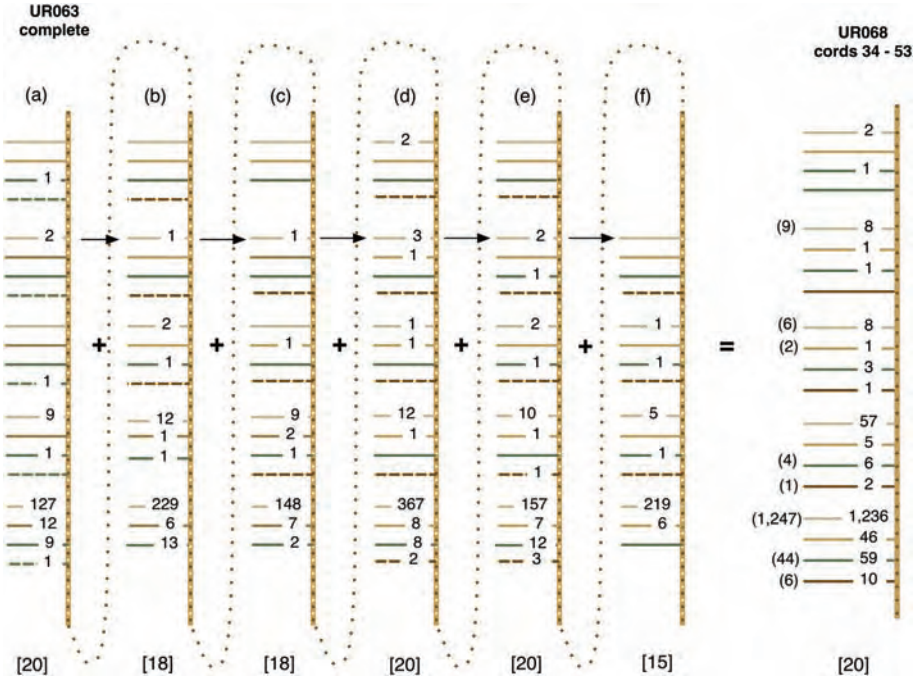


Fig. 2. Numerical and color correlations between khipu UR63 and the central section of UR68.

the right subunits. Except for the introductory segments, all strings on level III are involved in the summation relationship.

It appears that the original structure of the Puruchuco accounting hierarchy contained six paired khipu on level I, whose values were summed to produce those on the three subunits of the three khipu on level II, whose subunits in turn were summed and recorded on the two khipu on level III. Information was either being funneled and synthesized upward or subdivided and distributed downward among the three levels of khipu.

We assume that the Puruchuco accounting hierarchy was a set of records for use both within and outside the administrative center. Khipu on level III could represent either a set of instructions issued to the lord of Puruchuco from the provincial governor or reports on local Puruchuco resources to be sent to the provincial governor. In either of these scenarios, one of the requirements would have been that the khipu bear an indication of their destination or origination. If numerous khipu were coming into a central archive for storage or were being dispersed from that archive to disparate places, it would have been helpful, if not essential, to have place identifiers encoded within each khipu. We suggest that the introductory segments on level II and III khipu represented just such identity labels.

The numerical values knotted onto strings within the introductory segments on level II and III khipu all contain arrangements of just three figure-eight knots tied onto three separate strings. Figure-eight knots on khipu normally signify the numerical value one. We hypothesize that the arrangement of three figure-eight knots at the start of these khipu represented the place identifier, or toponym, "Puruchuco." We suggest that any khipu moving within the state administrative system bearing an initial arrangement of three figure-eight knots would have been immediately recognizable to Inka administrators as an account pertaining to the palace of Puruchuco.

Why don't level I khipu bear introductory segments? Perhaps UR63 and UR73 were not intended to travel away from Puruchuco; instead, they may have been local accounts, drawn up by the resident khipukamayuq for accounting purposes within the palace. If the seven khipu in Fig. 1 register demands for service received from outside Puruchuco, meaning that if the relation among them is one of partition, then the level I khipu would have represented the reorganization of the mandate from outside in relation to the availability of resources at the local level. In this scheme, level I khipu would have pertained only to local accounting matters, and it would have been unnecessary to attach the place identifier. However, if the overall relationship is one of summation, and these khipu were prepared as a report on local conditions for dispatch outside Puruchuco, then level I khipu would represent the raw tables of local information that served as the foundation for constructing level II and III khipu. Level III khipu, the summary reports, would have been sent to a distant administrative center.

We suggest that khipu may have contrasting number qualities depending on whether they represented instructions coming from the state administration to a local accounting center or were records produced within a local accounting center with regard to existing community resources. In the first circumstance, we suspect that khipu values would have tended to be even decimal values or calculations of values in standard proportional shares. If a khipu account was compiled from within some local administrative center to be sent upward to higher level officials, counts of resources could be expected to have reflected the vagaries of the natural distribution of items in society. Such numbers are less likely to be whole and rounded or perfectly proportional.

We believe that the Puruchuco archive is the first known example indicating how information moved both up and down the Inka administrative hierarchy. There is insufficient evidence to determine whether the khipu are related through data partition or summation; however, careful study of the Puruchuco and other khipu archives may provide the foothold needed for addressing the most difficult question facing students of the Inka khipu: How did the khipu-keepers of the Inka administrative system record the identities of objects—people, animals, produce, manufactured goods, etc.—in the three-dimensional forms of their knotted-string records (*7*)?

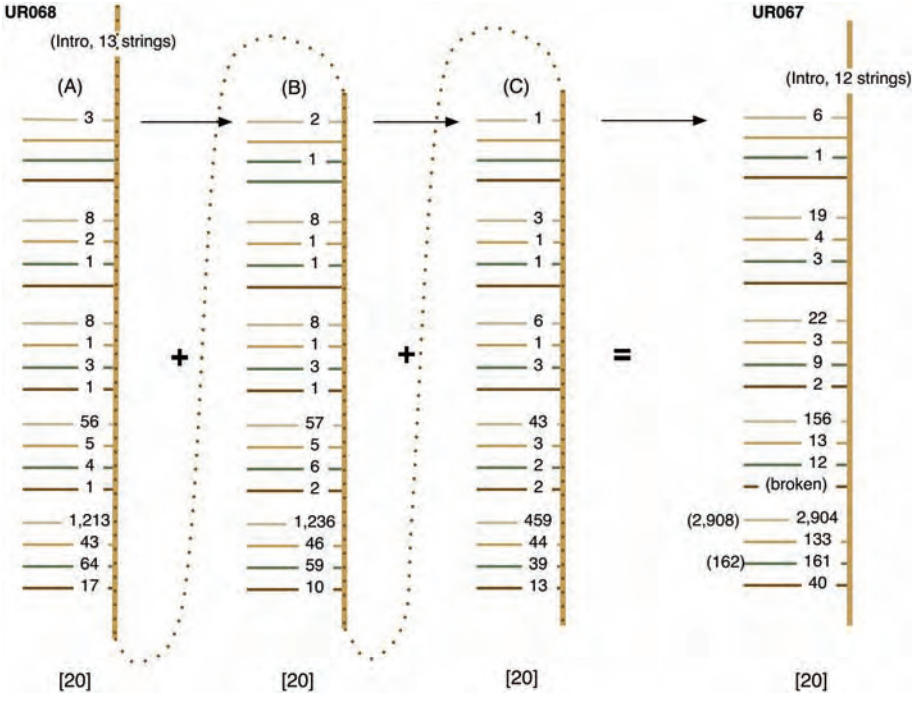


Fig. 3. Numerical and color correlations between khipu UR68 and UR67.

References and Notes

1. G. de la Vega, El Inca, *Royal Commentaries of the Incas* (Univ. of Texas Press, Austin, TX, 1966).
2. C. Mackey, thesis, University of California, Berkeley (1970).
3. Thanks to Julio Tello Solis for his transcription (10 July 2004) of the notes from the excavation field reports at Puruchuco, here translated by Urton: "9 August, 1956—The work consisted, as over the past three days, in removing 'fill,' or dirt from the upper part of sector B to fill a pit in Platform A … In the zone of extraction (Sector B, upper part) of the fill the workman Lizama encountered a narrow-necked urn (cantaro), semi-ovoid in form, covered with soot (hollín) and with an applique on the outer body in the form of a serpent; its [i.e., the urn's] mouth was covered by a small lagenaria [bottle gourd]; in the interior there were found 10 khipu of regular size, 3 of which had red/orange/yellow tassels, 11 medium sized ones, and several loose pendant strings, all in a good state of preservation."
4. Sample 9 was in the Puruchuco museum when Carol Mackey studied this collection in the 1960s. When we restudied the Puruchuco khipu archive in the summer of 2004, 9 was no longer in the collection. Museum personnel could not tell us what had become of this sample.
5. C. Radicati de Primeglio, *La 'Seriación' como posible Clave para Descifrar los Quipus Extranumerales* (Biblioteca de la Sociedad Peruana de Historia, Lima, Peru, 1964).
6. F. Salomon, *The Cord Keepers: Khipus and Cultural Life in a Peruvian Village* (Duke Univ. Press, Durham, NC, 2004), pp. 252–255.
7. The Khipu Database project, located in the Department of Anthropology, Harvard University, is described fully on the project Web site at http://khipukamayuq.fas.harvard.edu/.

8. We thank L. F. Villacorta Ostolaza and the staff of the Museo de Sitio Puruchuco–Arturo Jiménez Borja: Bullón, Díaz, and Solis. We thank NSF (grant BCS-0408324); the Dumbarton Oaks Foundation; and the Faculty of Arts and Sciences, Harvard University, for support; and the John D. and Catherine T. MacArthur Foundation for G.U.'s MacArthur Fellowship.

11 April 2005; accepted 13 July 2005

Nature of Phosphorus Limitation in the Ultraoligotrophic Eastern Mediterranean

T. F. Thingstad,[1]* M. D. Krom,[2] R. F. C. Mantoura,[3,4] G. A. F. Flaten,[1] S. Groom,[3] B. Herut,[5] N. Kress,[5] C. S. Law,[3,6] A. Pasternak,[7] P. Pitta,[8] S. Psarra,[8] F. Rassoulzadegan,[9] T. Tanaka,[1,9] A. Tselepides,[8] P. Wassmann,[7] E. M. S. Woodward,[3] C. Wexels Riser,[7] G. Zodiatis,[10] T. Zohary[11]

Phosphate addition to surface waters of the ultraoligotrophic, phosphorus-starved eastern Mediterranean in a Lagrangian experiment caused unexpected ecosystem responses. The system exhibited a decline in chlorophyll and an increase in bacterial production and copepod egg abundance. Although nitrogen and phosphorus colimitation hindered phytoplankton growth, phosphorous may have been transferred through the microbial food web to copepods via two, not mutually exclusive, pathways: (i) bypass of the phytoplankton compartment by phosphorus uptake in heterotrophic bacteria and (ii) tunnelling, whereby phosphate luxury consumption rapidly shifts the stoichiometric composition of copepod prey. Copepods may thus be coupled to lower trophic levels through interactions not usually considered.

The eastern Mediterranean Sea is a large body of ultraoligotrophic (*1*) water where both nutrient stoichiometry (*2*), bioassays (*3–6*), and P-cycle studies (*7*) have suggested P limitation of biological production. We performed the Cycling of Phosphorus in the Eastern Mediterranean (CYCLOPS) experiment in the eastern Mediterranean near the center (33.3°N, 32.3°E) of the quasi-stable, warm-core Cyprus Eddy (Fig. 1). The area was studied for 4 days before the release of phosphate on 17 May 2002. Diluted phosphoric acid was added to an ~16-km^2 patch to a concentration of ~110 nM (*8*). Vertical spreading was restricted by the seasonal pycnocline at 16 m. The patch, defined by the simultaneous addition of an SF_6 inert tracer, was monitored for 9 days with biological and chemical investigations at stations located in the center of the patch (IN stations). On every other day, the ship moved to a minimum distance of 15 km from the patch center for vertical profiling at an OUT station.

Pre-experiment investigations confirmed the ultraoligotrophic status of the surface waters, which had chlorophyll concentrations of 18 ± 1 μg of chl m^{-3} (Fig. 2A) and a daily primary productivity of 29 μmol of C m^{-3} day^{-1}. Pigment analysis of size-fractionated material indicated that most (~65%) of the chlorophyll a (chl-a) was in the picoplankton (<2 μm) fraction with a large (~40% of total chl-a) cyanobacterial component.

After deep mixing in the winter months, dissolved phosphate in the surface waters remained below our detection limits, whereas there is typically a small residual (300 to 500 nM) of nitrate (*2*, *9*). This nitrate was isotopically heavy, which is characteristic of a phytoplankton bloom that has ceased because of P limitation (*10*).

By May, both phosphate and nitrate concentrations were below the detection limit (~2 nM) of the nanomolar technique used (*8*). Using a titrimetric bioassay method based on alkaline phosphatase activity, we estimated the system to contain ~230 ± 60 nM residual N (*5*), which suggests that a large fraction of the residual N had been retained in the system in a form available to at least some components of the biota. Ammonium, determined to 60 to 80 nM N, may have accounted for about one-third of this excess N; the form of the remaining is unknown. The surface waters also contained 65 to 100 μM dissolved organic carbon (DOC), 3 to 11 μM dissolved organic nitrogen (DON), and ~50 nM ultraviolet-oxidizable dissolved organic phosphorus (DOP). All physiological and biogeochemical indicators examined pointed to a P-starved system: The DON/DOP molar ratio was 60:1 to 200:1, the PON/POP ratio was 30:1, and there was alkaline phosphatase activity in surface waters outside the patch (*5*). Despite the low biomass (7 to 10 nM POP) (Fig. 2B), turnover time for orthophosphate was low for a marine system (2 to 4 hours) (Fig. 3). Combining biomass and turnover time gave us a conservative estimate of the affinity constant for orthophosphate uptake of ~0.08 liter nM-P^{-1} h^{-1} (supporting online text), consistent with what would be expected for orthophosphate uptake limited by molecular diffusion in water (*11*). Maximum potential for orthophosphate uptake was determined as 1.6 ± 0.2 μmol P m^{-3} h^{-1}, which is about one order of magnitude above the P demand that can be calculated from a stoichiometric conversion of bacterial and primary carbon production (supporting online text).

From the SF_6 tracer (Fig. 3), we estimated a lateral diffusion coefficient for the patch of

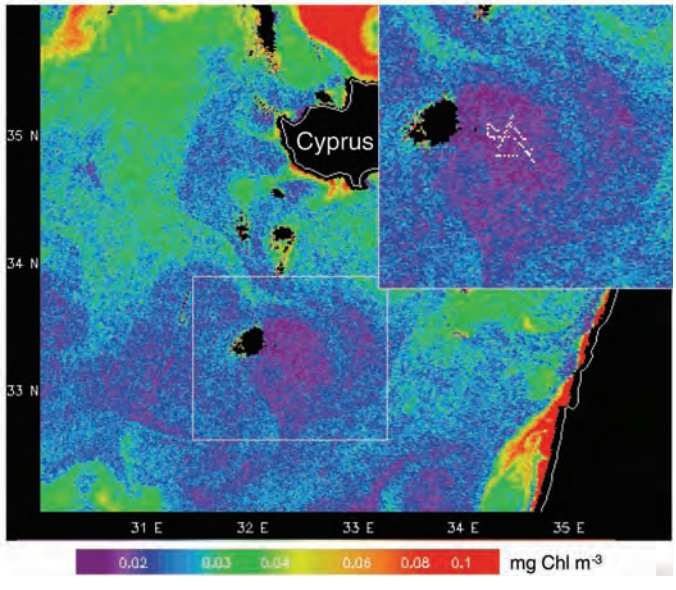


Fig. 1. SeaWiFS satellite chl-a image of the eastern Mediterranean on 25 May 2002, showing the experimental area (inset) in the center of the oligotrophic Cyprus Gyre (the cruise track is in white). The color scale is in mg of chl-a m^{-3}. The black patch to the west of the experimental area is a cloud.

[1]Department of Biology, University of Bergen, Bergen, Norway. [2]School of Earth and Environmental Sciences and Earth and Biosphere Institute, Leeds University, Leeds, UK. [3]Plymouth Marine Laboratory, Plymouth, UK. [4]Marine Environment Laboratory, International Atomic Energy Agency, Monaco. [5]Israel Oceanographic and Limnological Research (IOLR), National Institute of Oceanography, Tel Shikmona, Haifa, Israel. [6]National Institute of Water and Atmospheric Research, Wellington, New Zealand. [7]Norwegian College of Fishery Science, University of Tromsø, Tromsø, Norway. [8]Hellenic Centre for Marine Research, Heraklion, Crete, Greece. [9]Station Zoologique, Villefranche-sur-mer, France. [10]Oceanography Centre, University of Cyprus, Nicosia, Cyprus. [11]IOLR, Kinneret Limnological Laboratory, Tiberias, Israel.

*To whom correspondence should be addressed. E-mail: frede.thingstad@bio.uib.no

21 ± 2 m^{-2} s^{-1}, corresponding to an initial dilution rate of the P-fertilized patch of ~1.0 to 1.2 day^{-1}, decreasing rapidly to ~0.1 day^{-1} after day 6. This corresponds to a theoretical dilution of the added P to a nominal level of ~15 nM after ~7 to 8 days. Given a Redfield N:P ratio of 16:1 (molar), the estimated excess N of 230 nM N would correspond to a requirement of ~14 nM P for its consumption. One would thus expect the system, through N import by physical mixing, to return to its original P-limited state after ~1 week. This scenario is confirmed by the measured orthophosphate turnover time, which increased more than 20-fold after *P* addition. By 1 week after addition, turnover time had returned to <10 hours (Fig. 3), and orthophosphate concentration had returned to values below our detection limit (~2 nM).

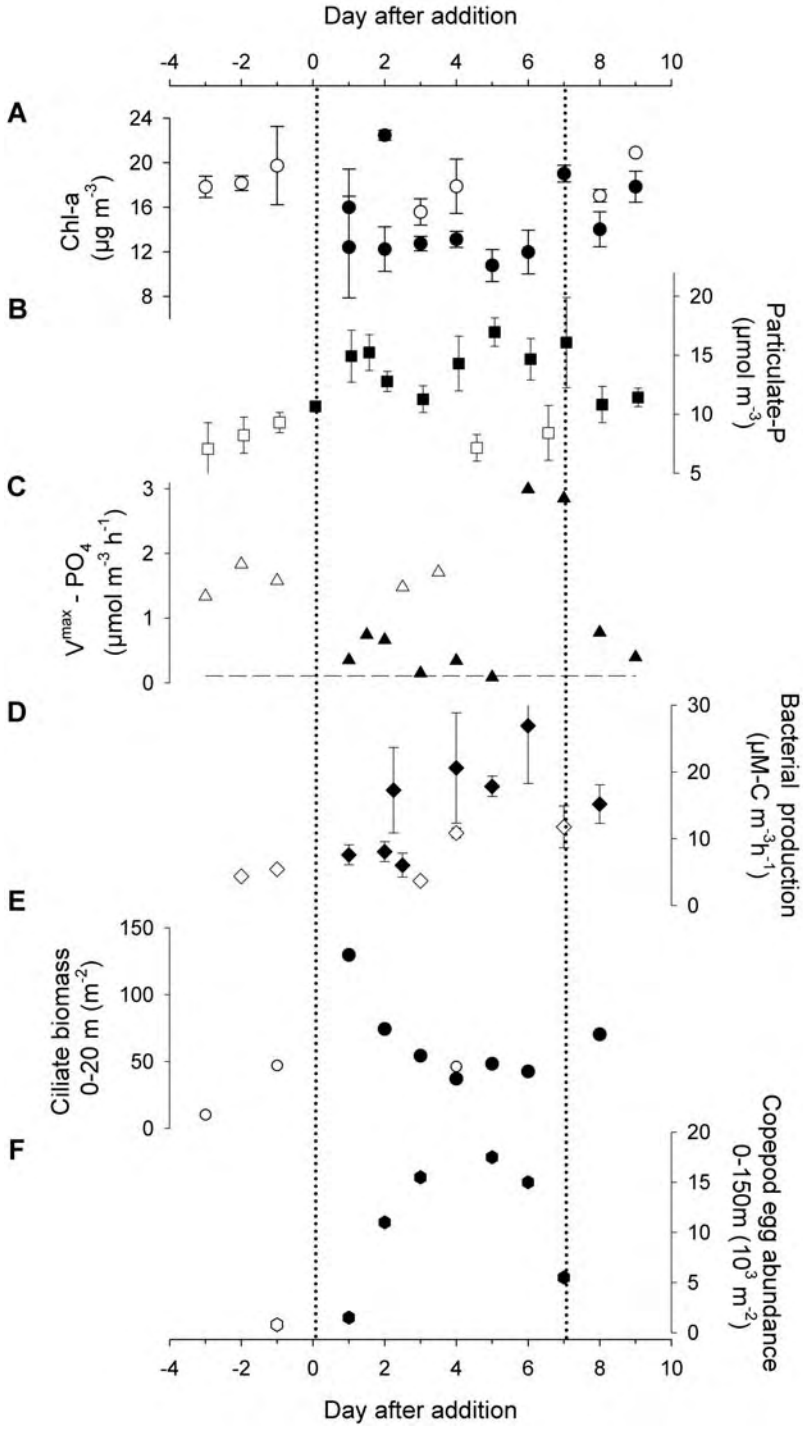


Fig. 2. Biological responses in the top layer: (**A**) High-pressure liquid chromatography–measured chl-a (circles), (**B**) particulate P (squares), (**C**) maximum uptake potential for orthophosphate (triangles, V_{max}), and (**D**) bacterial production (diamonds), all averaged over depths <20 m. Error bars show the SE of samples (generally four) taken within the 0- to 16-m layer and thus contain both experimental error and natural variations within the top layer. (**E**) Ciliate biomass integrated over 0 to 20 m (circles) and (**F**) abundance of copepod eggs integrated over 0 to 150 m (hexagons). Open symbols indicate samples taken before or outside the patch; solid symbols, samples taken inside the patch.

If all of the assayed 230 nM excess N (*5*) was assimilated by *Synechococcus* sp., and assuming a N:chl-a ratio of 0.31 nmol N per ng of chl-a for P-replete *Synechococcus* sp. (*12*, *13*), this would, in the absence of dilution and grazing, correspond to a potential for an ~40-fold increase in chlorophyll (from 18 to 741 µg of chl-a m^{-3}). Instead, the system responded to the addition of phosphate with a decrease of chl-a in the patch of ~40% (Fig. 2A) [one-way analysis of variance (ANOVA) comparing samples inside the patch with samples taken outside the patch and before the experiment, $P < 0.05$], reaching a minimum of 11 µg of chl-a m^{-3} after 5 days. Chlorophyll levels returned to background levels after ~1 week. There were also decreasing trends in primary production and in phytoplankton growth rates, while picophytoplankton and smaller nanophytoplankton were reduced in numbers (by 49 and 65%, respectively) within the first 4 days of the experiment. With the precision obtainable in this type of experiment, precise estimates cannot be given for the contributions of growth (µ) and loss rate (γ) to the net change in chlorophyll. When corrected for dilution, a rough estimate suggests a post-addition net loss of $\mu - \gamma \approx -0.6$ day^{-1}, equally partitioned between a decrease in µ and an increase in γ (supporting online text).

In an on-deck microcosm experiment, ammonium was added or not to water collected inside or outside the P-fertilized patch. Ammonium addition induced a phytoplankton bloom in water collected inside the patch but not in water collected outside the patch (fig. S1) (supporting online text). The in situ finding that adding P alone leads to increased particulate P and bacterial production but not phytoplankton production was replicated in the microcosms. This demonstrated the N-limited state of phytoplankton after P addition and indicates that the assayed excess N was not available to phytoplankton over this time scale. The natural system, despite clear indications of P starvation and excess N, could thus be described as colimited by N and P for phytoplankton.

N_2 fixation has been proposed as an important mechanism driving the Mediterranean toward P limitation (*14*). If N_2 fixation played a role in our experiment, however, either it was insufficient to produce a positive net response in phytoplankton or it occurred in heterotrophs. In contrast to the negative trend in phytoplankton response, particulate P (Fig. 2B) had almost doubled already at the first

sampling after addition, concurrent with a drop in the maximum potential for orthophosphate uptake (Fig. 2C). This is the pattern that would be expected for phytoplankton (*15*) and heterotrophic bacteria (*16*) rapidly replenishing their internal cell quota for P and reducing their maximum uptake capacity to the level required for growth.

There was an increase in bacterial production (ANOVA, comparing measurements before and inside the patch, $P < 0.001$) (Fig. 2D). Bacterial abundance ranged from 0.6 to 1.4 × 10^{11} cells m^{-3}, with no clear trend between samples taken inside and outside the patch. We suggest that the nitrogen and organic C required to support this increased bacterial production may be from the DON and DOC pools, respectively.

Copepod egg abundance in the water column increased by more than one order of magnitude, a response visible after 2 days and peaking after 5 days (Fig. 2F). There may also have been a positive response in ciliate abundance peaking on the first day after addition (Fig. 2E) (day 1 was significantly different from days –3 and –1, and from days 5, 6, and 8, $P < 0.01$), then disappearing rapidly.

We propose two, not mutually exclusive, trophic pathways for transfer of the added P to copepods. One is a "trophic bypass" whereby added P bypasses the phytoplankton compartment through a predatory food chain from heterotrophic bacteria (Fig. 4), producing a concomitant increase in predation loss (γ) of small-celled phytoplankton. Support for such a mechanism can be found in our stimulated bacterial production (Fig. 2D). Another possible route for P transfer is by "trophic tunnelling," in which phosphate "disappears" through rapid luxury consumption into phytoplankton and bacteria, thus changing the stoichiometry but not the abundance of prey organisms. If there are predator processes such as ciliate DNA synthesis or copepod egg production that are P-limited, the added P would "reappear" as responses at the predator level much more rapidly than could be expected from models that transport P through nutrient-prey-predator successions in a food chain of organisms with fixed biomass stoichiometry.

The bypass mechanism would be in agreement with previous observations indicating P limitation of heterotrophic bacteria in the eastern Mediterranean (*4*), implying that there are components in the large DOC and DON pools accessible to bacteria. The relative importance of this bypass route would be enhanced if heterotrophic bacteria have better access than phytoplankton to the system's pool of excess N. A likely candidate would be dissolved combined amino acids, usually believed to be primarily accessible by heterotrophic bacteria. This would imply that there are mechanisms in the system converting any nitrate remaining after the winter phytoplankton bloom to DON in the summer, thus converting an apparently conventionally defined P-limited system to one where phytoplankton are colimited by P and N. Extreme situations with pure N limitation of autotrophs while bacteria are P-limited have recently been described in coastal wetlands (*17*).

However, the sequence where ciliate (Fig. 2E) and copepod (Fig. 2F) responses occur before any detectable increase in bacterial production (Fig. 2D) is difficult to reconcile with bypass as the only trophic transport of P. Because P acquired by luxury consumption cannot be translated into biomass production in the absence of available N, lack of access to the excess N pool would leave the phytoplankton with a high P cell quota. Estimating osmotroph C biomass to ~50% of total microbial C, the approximate doubling in total particulate P may have represented an ~4 times increase in the osmotroph cell quotas. The amount of P transported per prey into the predator food chain must have increased accordingly. It is thus reasonable to adapt the limnologist's hypothesis of P-starved mesozooplankton (*18*). A change in food quality (P content), rather than in food quantity (abundance), is thus suggested as the main signal initiating egg production in copepods and possibly also cell division in ciliates, shortly after the orthophosphate addition. If, in addition to the traditional luxury consumption, there was a surface adsorption of added phosphate to phytoplankton (*19*), this would add to the proposed tunnelling transport.

Implicit in this scenario is that heterotrophs use the time between nutrient pulses to store carbon (energy) and nitrogen, bacteria possibly in extracellular forms (such as DOC and DON) and zooplankton probably as somatic storage material. Both groups would then be able to respond once P becomes available, as orthophosphate or as P-enriched prey, respectively.

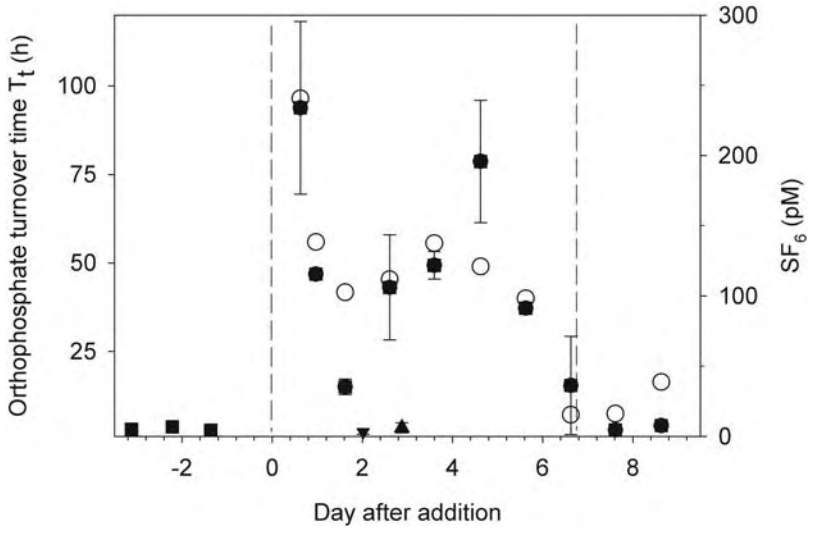


Fig. 3. SF_6 concentrations (○) and orthophosphate turnover time (T_t) before (■), inside (●), outside (▼), and below (▲) the patch. Vertical dotted lines indicate the experimental release and the estimated time of return of the system from N to P limitation. Error bars indicate SE of four samples taken inside the 0- to 16-m layer.

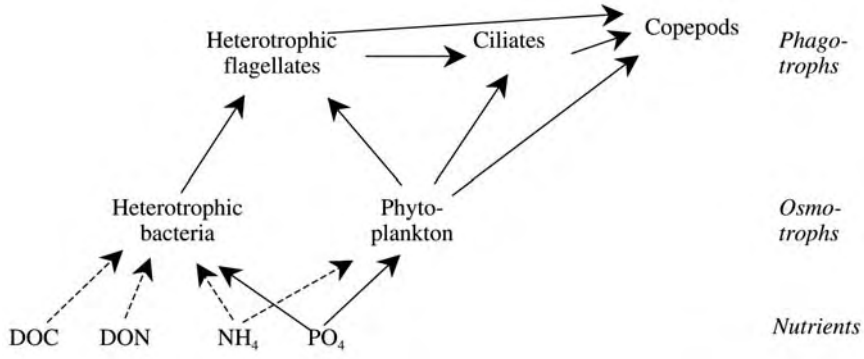


Fig. 4. Idealized model of the P flow (solid arrows) through the lower part of the pelagic food web. We suggest that the added P can be transported to the level of phagotroph predators either through bypass, in which heterotrophic bacteria can produce biomass because of access to parts of the DON pool, or by tunnelling, in which luxury consumption increases the cell quota of P in osmotrophs, thus changing the quality but not the quantity of prey for P-limited predators. Other potentially limiting nutrients are indicated by broken arrows. The remineralization process is omitted for clarity.

With a summer situation characterized by P-limited bacteria and grazers, whereas phytoplankton are closer to an N and P colimitation, nature of the P limitation in the eastern Mediterranean becomes both seasonally variable and organism-dependent. Although our interpretation requires a much more elaborate conceptual model of the pelagic food web than is needed to explain the more conspicuous response of a phytoplankton bloom observed when iron has been added in high nutrient, low chlorophyll (HNLC) areas, the interpretation introduces no fundamentally new biological mechanisms.

Within our proposed interpretation, the counterintuitive combination of a negative chlorophyll response with a positive zooplankton response is a function of the oligotrophy, the P limitation, and the form in which excess N is stored and is thus linked to particular biogeochemical features of the post-bloom eastern Mediterranean ecosystem. Flexible stoichiometry, alternative microbial food chains, and differential access to pools of the next limiting element may be sophistications that can be disregarded in some biogeochemical models. In the eastern Mediterranean with its pulsed atmospheric nutrient inputs from Saharan dust storms, such trophic "sophistications" may however be the key to central aspects of ecosystem functioning.

Copepods are believed to have a central role not only in bridging production in the microbial part of the food web to commercially interesting fish resources but also as producers of vertical export of C, N, and P to the ocean's interior. With a low nutrient, low chlorophyll state being the dominant mode in the world's surface oceans, an increasing awareness of P-limited oceanic regions (*20*–*24*), and an increasing trend toward a high N:P in atmospheric inputs (*25*), these results are likely to be relevant not only to the eastern Mediterranean but also to large areas of today's oceans and their response to global change.

References and Notes

1. A. C. Redfield, B. H. Ketchum, F. A. Richards, in *The Sea*, M. N. Hill, Ed. (Wiley, New York, 1963), vol. 2, pp. 26–77.
2. M. D. Krom, N. Kress, S. Brenner, L. I. Gordon, *Limnol. Oceanogr.* **36**, 424 (1991).
3. D. J. Bonin, M. C. Bonin, T. Berman, *Aquat. Sci.* **51**, 129 (1989).
4. T. Zohary, R. D. Robarts, *Limnol. Oceanogr.* **43**, 387 (1998).
5. T. F. Thingstad, R. F. C. Mantoura, *Limnol. Oceanogr. Methods* **3**, 94 (2005).
6. F. Van Wambeke, U. Christaki, A. Giannakourou, T. Moutin, K. Souvemerzoglou, *Microb. Ecol.* **43**, 119 (2002).
7. T. Moutin *et al.*, *Limnol. Oceanogr.* **47**, 1562 (2002).
8. Materials and methods are available as supporting material on *Science* Online.
9. N. Kress, B. Herut, *Deep-Sea Res. I* **48**, 2347 (2001).
10. U. Struck, K.-C. Emies, M. Voss, M. D. Krom, G. H. Rau, *Geochim. Cosmochim. Acta* **65**, 3249 (2001).
11. T. F. Thingstad, F. Rassoulzadegan, *Prog. Oceanogr.* **44**, 271 (1999).
12. S. Bertilsson, O. Berglund, D. M. Karl, S. W. Chisholm, *Limnol. Oceanogr.* **48**, 1721 (2003).
13. S. W. Jeffrey, H. S. MacTavish, W. C. Dunlap, M. Vesk, K. Groenewoud, *Mar. Ecol. Prog. Ser.* **189**, 35 (1999).
14. J. P. Béthoux, P. Morin, D. Ruiz-Pino, *Deep-Sea Res. II* **49**, 2007 (2002).
15. H. Ducobu, J. Huisman, R. R. Jonker, L. R. Mur, *J. Phycol.* **34**, 467 (1998).
16. O. Vadstein, *Aquat. Microb. Ecol.* **14**, 119 (1997).
17. P. V. Sundareshwar, J. T. Morris, E. K. Koepfler, B. Fornwalt, *Science* **299**, 563 (2003).
18. T. Andersen, J. J. Elser, D. O. Hessen, *Ecol. Lett.* **7**, 884 (2004).
19. S. A. Sanudo-Wilhelmy *et al.*, *Nature* **432**, 897 (2004).
20. J. Ammerman, R. Hood, D. Case, J. Cotner, *Eos* **84**, 165 (2003).
21. S. A. Sanudo-Wilhelmy *et al.*, *Nature* **411**, 66 (2001).
22. J. F. Wu, W. Sunda, E. A. Boyle, D. M. Karl, *Science* **289**, 759 (2000).
23. D. M. Karl *et al.*, *Deep-Sea Res. II* **48**, 1529 (2001).
24. M. M. Mills, C. Ridame, M. Davey, J. La Roche, R. J. Geider, *Nature* **429**, 292 (2004).
25. T. Jickells, in *Challenges of a Changing Earth*, W. Steffen, J. Jager, D. J. Carson, C. Bradshaw, Eds. (Springer-Verlag, New York, 2002), pp. 93–96.

26. We wish to thank the crew of research vessel *Aegaeo* for companionship and assistance, and P. Carbo for her invaluable help in administering the whole program. This project was funded by the European Union through project no. EVK3-CT-1999-00009, "CYCLOPS".

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5737/1068/DC1
Materials and Methods
SOM Text
Fig. S1
References and Notes

23 March 2005; accepted 12 July 2005
10.1126/science.1112632

Direct Control of Germline Stem Cell Division and Cyst Growth by Neural Insulin in *Drosophila*

Leesa LaFever and Daniela Drummond-Barbosa*

Stem cells reside in specialized niches that provide signals required for their maintenance and division. Tissue-extrinsic signals can also modify stem cell activity, although this is poorly understood. Here, we report that neural-derived *Drosophila* insulin-like peptides (DILPs) directly regulate germline stem cell division rate, demonstrating that signals mediating the ovarian response to nutritional input can modify stem cell activity in a niche-independent manner. We also reveal a crucial direct role of DILPs in controlling germline cyst growth and vitellogenesis.

Germline and somatic stem cells support oogenesis throughout adult life in *Drosophila* (*1*) (fig. S1). Germline stem cells (GSCs) reside within a specialized niche where they are exposed to a unique combination of signals required for stem cell function (*2*, *3*). However, GSCs are also controlled by tissue-extrinsic signals, such as *Drosophila* insulin-like peptides (DILPs), which mediate the ovarian response to nutrition (*4*). On a protein-rich diet, germline and somatic stem cells have high division rates, and their progeny exhibit high division and development rates. On a protein-poor diet or under reduced insulin signaling, rates of division and development are reduced, and progression through vitellogenesis is blocked (*4*). It remains unclear, however, how DILPs control the response of GSCs in coordination with their differentiating progeny and with follicle cells.

Department of Cell and Developmental Biology, Vanderbilt University Medical Center, 4120B Medical Research Building III, 465 21st Avenue South, Nashville, TN 37232–8240, USA.

*To whom correspondence should be addressed. E-mail: daniela.drummond-barbosa@vanderbilt.edu

In adult females, DILPs are produced in two clusters of medial neurosecretory cells in the brain (*5*), and stage 10 follicle cells express *dilp5* mRNA (*6*). Ablation of brain DILP-producing cells results in reduced egg production rates and a partial block in vitellogenesis (*6*). To examine the role of the brain DILP-producing cells in previtellogenic stages, we ablated them and measured follicle cell proliferation rates (*7*) (fig. S2). Females missing brain DILP-producing cells (ablated) have a severely impaired ability to up-regulate follicle cell proliferation in response to a protein-rich diet (Fig. 1). The rate of germline development is reduced in coordination with follicle cell divisions because no abnormalities are observed in previtellogenic egg chambers (*8*). Ablation of DILP-producing cells reduces the size of eclosed adults and delays development (*9*). Ablated females in which these developmental defects are rescued by an *hs-dilp2* transgene expressed during larval stages show a reduced follicle cell proliferation rate, comparable to that of nonrescued, ablated females (Table 1). Thus, the impaired response to a protein-rich diet is not a secondary consequence of the developmental defects. Moreover, the 2.3-fold delay caused by ablation of brain DILP-producing cells (*7*) is very similar to that caused by blocking reception of DILP signals by the germ line (below). This indicates that

the brain is the major source of DILPs that determine the rate of egg chamber development with little, if any, contribution from *dilp5*-expressing follicle cells.

To examine whether DILPs control the rate of germline development directly or indirectly, we created germline cysts unable to respond to DILPs, by inducing *Drosophila insulin receptor* (*dinr*) mutant clones using the flipase (FLP)/FLP-recognition target (FRT) technique (Fig. 2A). Germline cysts homozygous for $dinr^{339}$, a genetic null allele (*10*), had normal morphology and correct cell number (*8*); however, 83% of these cysts were developmentally delayed, showing markedly decreased size relative to neighboring wild-type egg chambers (*7*) (Fig. 2B and Table 2). Further quantification of these data showed a 2.4-fold delay in the development of $dinr^{339}$ cysts (*7*) (table S1). Similar results were obtained for germline cysts homozygous for $dinr^{E19}$ and $dinr^{353}$, which are viable hypomorphic alleles (*10*); 78% and 64% of $dinr^{E19}$ and $dinr^{353}$ cysts, respectively, were developmentally delayed (Fig. 2, C and D, and Table 2). These results reveal that *dinr* function is required cell autonomously for a normal rate of germline cyst development. Thus, the rate of cyst development is regulated by a DILP signal that is received directly.

Progression through vitellogenesis requires DILP signaling (*4*, *11*); however, it has been unclear whether this role is direct. Reduced juvenile hormone levels are present in homozygous viable *dinr* mutants, and the block in yolk uptake in these mutants can be partially bypassed by treatment with methoprene, a juvenile hormone analog (*12*), suggesting an indirect role for DILPs in promoting vitellogenesis. To specifically address whether direct activation of germline cysts by DILPs is required for vitellogenesis, we analyzed mosaic ovarioles in which the entire germ line was homozygous *dinr* mutant for the ability of their egg chambers to undergo vitellogenesis. All egg chambers containing $dinr^{339}$ (n = 11 egg chambers) or $dinr^{E19}$ (n = 18 egg chambers) homozygous mutant cysts failed to progress through vitellogenesis and degenerated (Fig. 2E). In the case of $dinr^{353}$, the allele with the higher level of *dinr* activity, only one out of six egg chambers containing homozygous mutant cysts failed to undergo vitellogenesis (*8*). These results suggest that the level of insulin signaling within the germ line controls vitellogenesis, revealing a direct role for DILPs in this process. Moreover, complete loss of *dinr* function in the germ line results in a complete block in vitellogenesis, whereas this block is partial upon ablation of brain DILP-producing cells (*6*). Thus, DILP5 expressed in stage 10 follicle cells likely signals in combination with brain DILPs to regulate vitellogenesis.

We next asked whether DILPs control GSC division rate directly by binding to receptors on their surface (a cell-autonomous requirement for *dinr* in GSCs) or indirectly by regulating signals produced by niche cells (a non–cell-autonomous requirement). We analyzed *dinr* mosaic ovarioles containing one wild-type and one mutant GSC (Fig. 2F) and counted the number of wild-type versus mutant cystoblasts and cysts present in their germaria. [The relative numbers of labeled versus unlabeled cysts was not affected by early germline cyst death (table S2).] Because each cystoblast or cyst corresponds to one GSC division (fig. S1C), the ratio of mutant to wild-type cystoblasts and cysts is a measure of their relative division rates (*7*). For $dinr^{339}$ homozygous mutant GSCs, we found a relative division rate of 0.31, whereas, for wild-type GSCs, it was 0.90 (Table 2). Similarly, the relative division rates of $dinr^{353}$ and $dinr^{E19}$ GSCs were 0.55 and 0.65, respectively (Table 2). Thus, *dinr* homozygous mutant GSCs divide more slowly than wild-type GSCs, and GSC division rate appears sensitive to the level of *dinr* activity. These results demonstrate that GSCs directly receive the DILP signal to regulate their division rate without mediation by the stem cell niche.

Germline and somatic cells respond to nutritional status in a coordinated manner; however, it is unclear whether somatic cells receive the DILP signal directly (a cell-autonomous role of *dinr* in follicle cell proliferation) as does the germ line, or indirectly through secondary signals (a non–cell-autonomous role). We measured the percentages of *dinr* mutant and control follicle cells in mosaic ovarioles carrying one wild-type and one *dinr* mutant somatic stem cell. If follicle cells receive the DILP signal directly, the reduced level of insulin signaling in *dinr* mutant follicle cells should result in lower rates of proliferation (i.e., fewer mutant than control follicle cells should be observed), whereas if they receive the signal

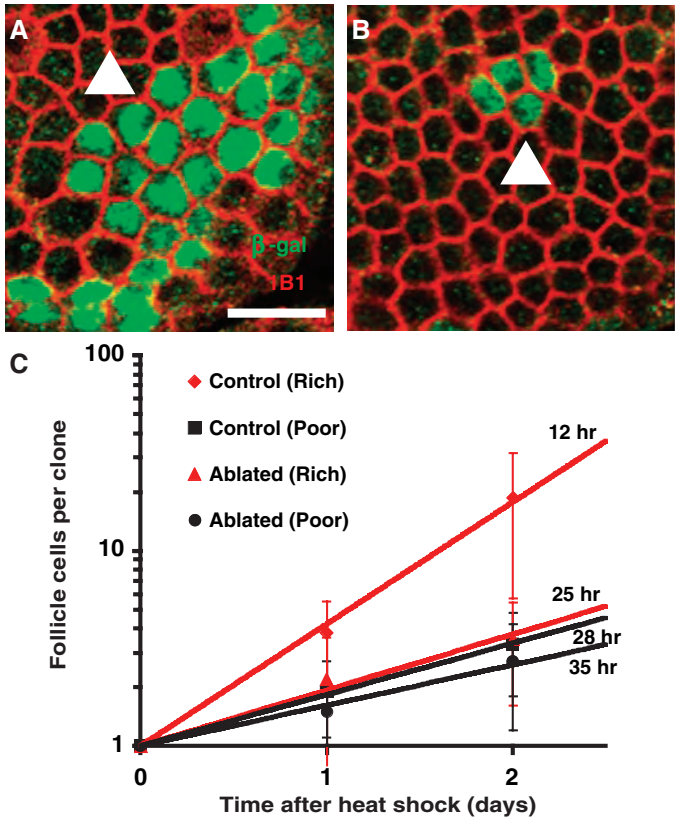


Fig. 1. Brain DILP-producing cells are required for the ovarian response to nutrition. At different times after heat shock, ovaries were dissected and immunostained for visualization of β-galactosidase (β-gal, green) and cell membranes (1B1, red). (**A** and **B**) Examples of egg chambers containing follicle cell clones (arrowheads) at 2 days after heat shock in (A) control or (B) ablated females on a rich food source are shown at the same magnification. (**C**) The number of follicle cells per clone was counted, and the average clone sizes in control or ablated females on rich or poor food sources were plotted. Bars show standard deviation. The responses of control and ablated females to a rich source are significantly different from each other ($P < 0.001$). Doubling times are shown next to corresponding curves. Scale bar, 10 μm.

Table 1. The impaired ovarian response to nutrition in females lacking brain DILP-producing cells is not due to developmental defects.

Strain (*7*)	Status of DILP-producing cells	Mitotic index
Control	Present	2.05% (9,080)*
Ablated	Absent	0.89% (11,631)†
Ablated, developmentally rescued	Absent	0.84% (12,962)

*The total number of follicle cells analyzed is shown in parentheses. †The mitotic indices for "ablated" versus "ablated, rescued development" females were not significantly different from each other, whereas these numbers were significantly different from the mitotic index for control females ($P < 0.001$).

indirectly, the proliferation rates should be similar. In *dinr*^{E19} mosaic ovarioles, 51% of follicle cells were wild-type and 49% were mutant (n = 7113 follicle cells), indicating similar proliferation rates (Fig. 2G). *dinr* mutant follicle cells appeared to enter the endoreplicative cycle normally, but pycnotic (degenerating) nuclei and cell death were observed within *dinr*^{E19} and *dinr*^{339} mutant follicle cell clones starting at stage 8 (*8*). These results reveal that although a reduction in *dinr* activity delays germline cyst development cell autonomously, it does not cause a cell-autonomous reduction in follicle cell proliferation rate. Furthermore, in ovarioles carrying a fully *dinr* mutant germ line (Fig. 2E), excess follicle cells were not observed, showing that proliferation of surrounding wild-type follicle cells remains coordinated with germline growth. These results suggest that follicle cells respond indirectly to increased DILP levels through a secondary signal from the germ line. Similar degrees of coordination between germ line and soma have been observed in the presence of developmentally delayed *dMyc* mutant germline clones (*13*).

These data demonstrate that tissue-extrinsic DILP signals can directly modify GSC proliferative activity, acting in parallel to signals from their niche. We also provide evidence that, in addition to its previously reported indirect roles in *Drosophila* and mammals through secondary hormonal signals (*12*, *14*), insulin signaling plays a crucial direct role during *Drosophila* oogenesis in regulating not only GSC division rate but also germline cyst development rate and progression through vitellogenesis. Insulin may, therefore, have important direct roles in mammalian oogenesis. Finally, our data suggest that the coordinated response of germline and somatic cells to nutrition involves communication between these tissues. These results have broad significance, in light of the long-known effects of nutrition on human fertility (*15*) and of the high degree of conservation of insulin signaling functions (*16*, *17*).

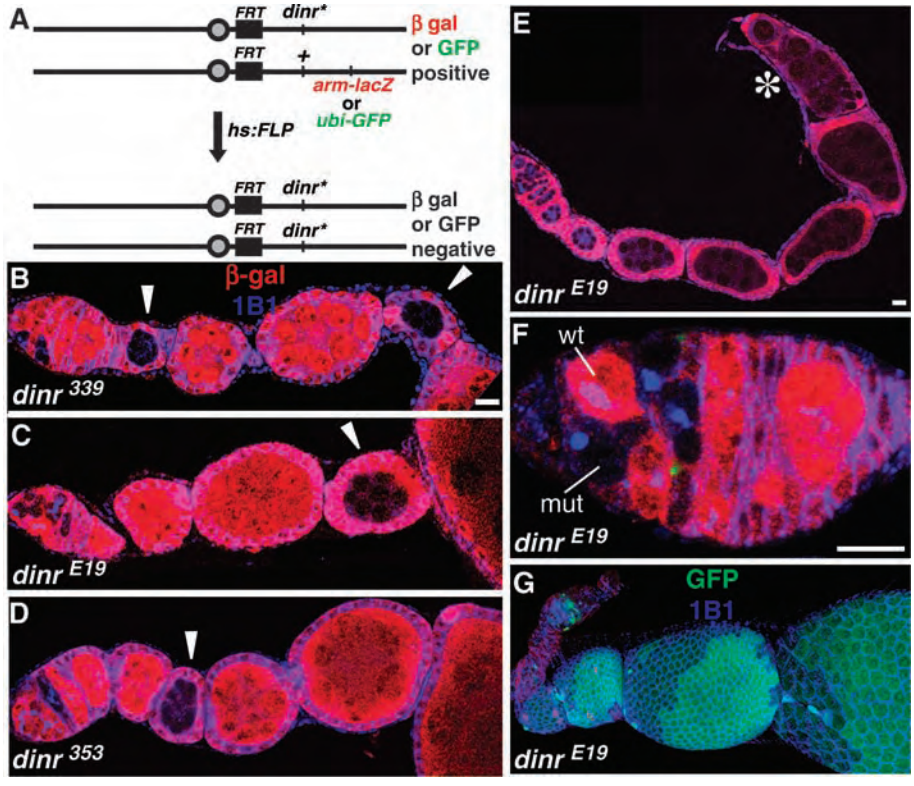


Fig. 2. *dinr* is cell-autonomously required in the germ line for normal rates of GSC division, cyst development, and progression through vitellogenesis. (**A**) The *FLP/FRT* technique (*18*) was used to generate mosaic ovarioles containing *dinr* homozygous mutant cells in the context of nearby heterozygous tissue. Flies carrying a wild-type *dinr* allele (+) in *trans* to a mutant or control *dinr* allele (*dinr**) were heat-shocked (hs) to induce FLP-mediated recombination between the *FRT* sites. Stem cell–derived *dinr* homozygous mutant clones were recognized by the absence of β-gal or green fluorescent protein (GFP) markers. (**B** to **D**) Homozygous (B) *dinr*^{339}, (C) *dinr*^{E19}, and (D) *dinr*^{353} germline cysts (arrowheads) are developmentally delayed relative to heterozygous control cysts. (B), (C), and (D) are shown at the same magnification. (**E**) Ovariole containing a germ line that is a fully *dinr*^{E19} homozygous mutant (and surrounded by wild-type follicle cells), showing a degenerating egg chamber (asterisk) that has failed to undergo vitellogenesis. (**F**) Example of mosaic germaria used to determine the relative division rate of *dinr*^{E19} homozygous mutant (mut) and control (wt) GSCs (Table 2). (**G**) Mosaic ovariole showing similar numbers of homozygous *dinr*^{E19} mutant (GFP-negative) and control (GFP-positive) follicle cells. 1B1 antibodies (blue) highlight cell membranes, whereas antibodies to β-gal (red) or GFP (green) mark control cysts. (E) and (G) are shown at the same magnification. Scale bars, 10 μm.

Table 2. DILPs directly regulate the rates of germline cyst development and of GSC division.

Strain	Percentage of marker-negative cysts showing delayed development (*7*)	Relative GSC division rate‡
Control	0% (96)*	0.90 (419)§
dinr^{339}	83% (118)†	0.31 (693)‖
dinr^{E19}	77% (92)	0.65 (321)
dinr^{353}	64% (123)	0.55 (461)

*The total number of marker-negative germline cysts analyzed is shown in parentheses. †The percentage of delayed cysts for each of the *dinr* mutants was statistically different from that of the control ($P < 0.001$ for each case). ‡Number of marker-negative cystoblasts and cysts divided by number of marker-positive cystoblasts and cysts (*7*). §The total number of cystoblasts and cysts counted is shown in parentheses. ‖The relative division rate for each of the *dinr* mutants was significantly different from that of controls ($P < 0.001$ for *dinr*^{353} and *dinr*^{339}, and $P < 0.04$ for *dinr*^{E19}).

References and Notes

1. A. C. Spradling *et al.*, *Cold Spring Harb. Symp. Quant. Biol.* **62**, 25 (1997).
2. A. Spradling, D. Drummond-Barbosa, T. Kai, *Nature* **414**, 98 (2001).
3. T. Xie, A. C. Spradling, *Cell* **94**, 251 (1998).
4. D. Drummond-Barbosa, A. C. Spradling, *Dev. Biol.* **231**, 265 (2001).
5. C. Cao, M. R. Brown, *Cell Tissue Res.* **304**, 317 (2001).
6. T. Ikeya, M. Galic, P. Belawat, K. Nairz, E. Hafen, *Curr. Biol.* **12**, 1293 (2002).
7. Materials and methods are available as supporting material on *Science* Online.
8. L. LaFever, D. Drummond-Barbosa, unpublished data.
9. E. J. Rulifson, S. K. Kim, R. Nusse, *Science* **296**, 1118 (2002).
10. W. Brogiolo *et al.*, *Curr. Biol.* **11**, 213 (2001).
11. C. Chen, J. Jack, R. S. Garofalo, *Endocrinology* **137**, 846 (1996).
12. M. Tatar *et al.*, *Science* **292**, 107 (2001).
13. J. Z. Maines, L. M. Stevens, X. Tong, D. Stein, *Development* **131**, 775 (2004).
14. J. C. Bruning *et al.*, *Science* **289**, 2122 (2000).
15. A. Wynn, M. Wynn, *Nutr. Health* **9**, 43 (1993).
16. R. S. Garofalo, *Trends Endocrinol. Metab.* **13**, 156 (2002).
17. D. C. Goberdhan, C. Wilson, *Differentiation* **71**, 375 (2003).
18. T. Xu, G. M. Rubin, *Development* **117**, 1223 (1993).

19. We thank E. Hafen, E. Rulifson, and the Bloomington Stock Center for fly stocks and D. Greenstein, L. Lee, B. Appel, and R. Cox for valuable comments on this manuscript. Supported by Development Funds from the Vanderbilt University School of Medicine, by American Cancer Society grant no. IRG-58-009-47, and by National Institutes of Health grant no. GM 069875.

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5737/1071/DC1
Materials and Methods
Figs. S1 and S2
Tables S1 and S2

23 February 2005; accepted 23 June 2005
10.1126/science.1111410

TAZ, a Transcriptional Modulator of Mesenchymal Stem Cell Differentiation

Jeong-Ho Hong,[1] Eun Sook Hwang,[3] Michael T. McManus,[1] Adam Amsterdam,[1] Yu Tian,[4] Ralitsa Kalmukova,[1] Elisabetta Mueller,[5] Thomas Benjamin,[4] Bruce M. Spiegelman,[5] Phillip A. Sharp,[1] Nancy Hopkins,[1] Michael B. Yaffe[1,2]*

Mesenchymal stem cells (MSCs) are a pluripotent cell type that can differentiate into several distinct lineages. Two key transcription factors, Runx2 and peroxisome proliferator–activated receptor γ (PPARγ), drive MSCs to differentiate into either osteoblasts or adipocytes, respectively. How these two transcription factors are regulated in order to specify these alternate cell fates remains a pivotal question. Here we report that a 14-3-3–binding protein, TAZ (transcriptional coactivator with PDZ-binding motif), coactivates Runx2-dependent gene transcription while repressing PPARγ-dependent gene transcription. By modulating TAZ expression in model cell lines, mouse embryonic fibroblasts, and primary MSCs in culture and in zebrafish in vivo, we observed alterations in osteogenic versus adipogenic potential. These results indicate that TAZ functions as a molecular rheostat that modulates MSC differentiation.

Pluripotent MSCs can differentiate into several distinct cell types, including osteoblasts and adipocytes (*1*, *2*). Two key transcription factors, Runx2 (also called Cbfa1 or Pebp2αA) and PPARγ, drive MSCs to differentiate into either osteoblasts or adipocytes, respectively (*3*–*7*), and the differentiation of each lineage appears to be mutually exclusive and transcriptionally controlled (*8*, *9*). How these two transcription factors are regulated in order to specify these alternate cell fates is unknown.

The balance between MSC osteoblast and adipocyte differentiation is disrupted in various human diseases. For example, decreased bone formation accompanied by an increase in bone marrow adipogenesis occurs with aging and immobility or following corticosteroid use (*10*–*14*), whereas increased bone formation is observed in patients with progressive osseous hyperplasia who form heterotopic bone within their adipose tissue (*15*). In addition, MSCs are currently being explored as vehicles for cell-based skeletal therapies (*16*, *17*). Therefore, investigating the

[1]Center for Cancer Research, Department of Biology and [2]Division of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, E18-580, Cambridge, MA 02139, USA. [3]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Harvard Medical School, Boston, MA 02115, USA. [4]Department of Pathology, Harvard Medical School, 77 Louis Pasteur Avenue, Boston, MA 02115, USA. [5]Dana-Farber Cancer Institute and the Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA.

*To whom correspondence should be addressed. E-mail: myaffe@mit.edu

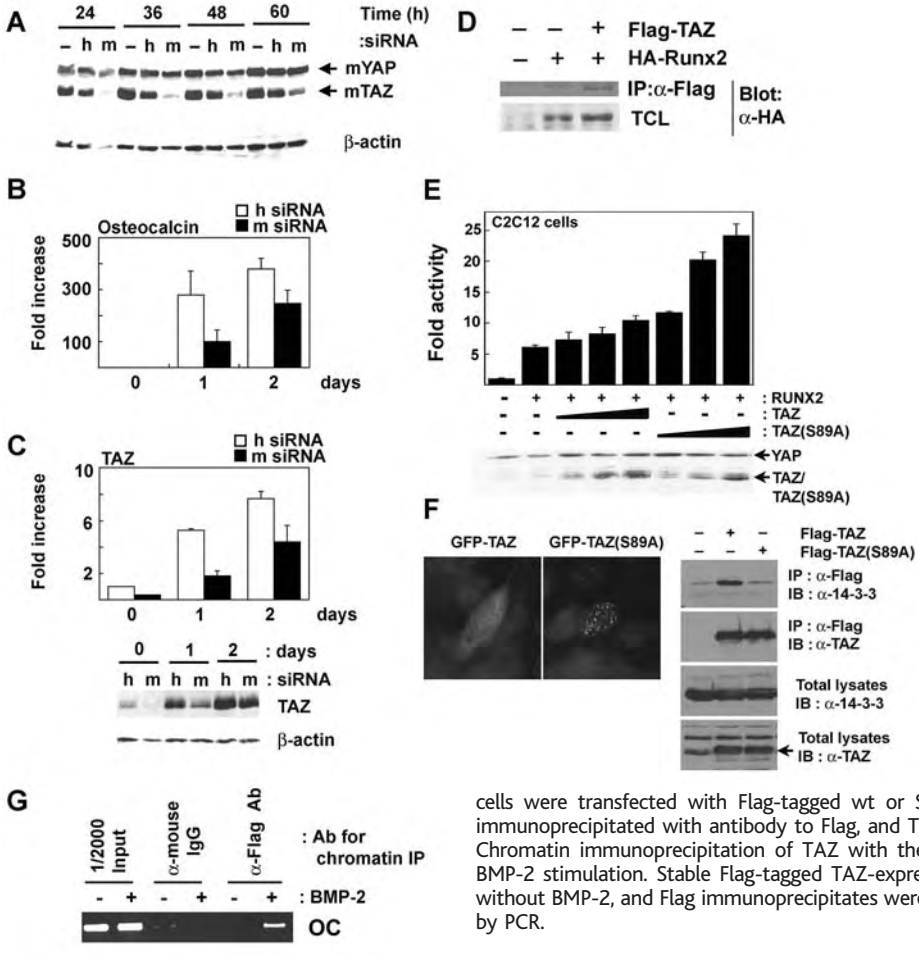


Fig. 1. TAZ is a Runx2 transcriptional coactivator for osteocalcin expression. (**A**) Down-regulation of TAZ by siRNA. Murine (m) and human (h) TAZ siRNA duplex oligonucleotides were transfected into murine C2C12 cells, and lysates were analyzed by immunoblotting (*18*). (**B**) Decreased octeocalcin expression after down-regulation of endogenous TAZ. Total RNA was prepared and osteocalcin gene expression at the indicated times after BMP-2 treatment was quantified by using real-time polymerase chain reaction (PCR) and normalized to β-actin expression. (**C**) TAZ expression in response to BMP-2. RNA and cell lysates from (B) were analyzed for TAZ mRNA expression by real-time PCR (upper panel) and for protein expression by Western blot analysis (lower panel). (**D**) Direct interaction between TAZ and Runx2. 293T cells were transfected with Flag-tagged TAZ, HA-tagged Runx2, or both, and total cell lysates (TCL) were analyzed by coimmunoprecipitation. (**E**) Stimulation of Runx2-driven gene expression by TAZ. Luciferase reporter activity from a construct containing six copies of the Runx2-binding site in the osteocalcin promoter was measured in cell lysates 24 hours after transfection and normalized to β-galactosidase activity. Below is a Western blot analysis of TAZ (wild-type or S89A) and YAP levels in the lysates. (**F**) 14-3-3 binding and nuclear localization of TAZ S89A mutant protein. (Left) Green fluorescent protein (GFP)–tagged TAZ wild-type and S89A localization in transfected C2C12 cells. (Right) C2C12 cells were transfected with Flag-tagged wt or S89A mutant TAZ. Protein from total lysates was immunoprecipitated with antibody to Flag, and TAZ or 14-3-3 was detected by immunoblotting. (**G**) Chromatin immunoprecipitation of TAZ with the endogenous osteocalcin promoter in response to BMP-2 stimulation. Stable Flag-tagged TAZ-expressing C2C12 cells were treated for 3 days with or without BMP-2, and Flag immunoprecipitates were analyzed for osteocalcin (OC) promoter occupancy by PCR.

mechanisms that fine-tune the balance between MSC osteoblast and adipocyte differentiation is likely to be of medical importance.

TAZ is a paralog (related molecule arising in a single species through gene duplication) of Yes-associated protein (YAP) that we identified in a proteomic screen for 14-3-3–binding proteins (*18*). The function of TAZ is unknown, although it can act as a transcriptional coactivator when overexpressed (*18*, *19*). TAZ and YAP contain a WW domain that binds to Pro-Pro-X-Tyr motifs (where X is any amino acid) and a coiled-coil C-terminal domain that recruits core components of the transcriptional machinery (*18*, *20*). A number of transcription factors including Runx2 and PPARγ contain Pro-Pro-X-Tyr motifs within their activation domains. Thus, interaction of TAZ with these and other transcription factors might be involved in mediating their transcriptional effects.

To investigate the function of TAZ, we designed small interfering RNAs (siRNAs) against the murine (m) or human (h) TAZ isoforms (fig. S1). These siRNAs differ at two base pairs and function as specific and nonspecific control siRNAs in mouse and human cell culture, respectively. The murine-specific TAZ siRNA, but not the human-specific siRNA, decreased the abundance of TAZ protein within 24 hours when transfected into murine myoblast C2C12 cells (Fig. 1A). Neither siRNA had any effect on YAP expression. Treatment of C2C12 cells with bone morphogenic protein–2 (BMP-2) induces osteoblast differentiation as evidenced by an ~400-fold increase in Runx2-dependent expression of the osteocalcin gene, a late marker of osteoblast development (*21*, *22*). Transfection of C2C12 cells with murine TAZ siRNA for 24 hours before exposure to BMP-2 inhibited osteocalcin gene expression relative to the effect of the human TAZ siRNA (Fig. 1B). BMP-2 treatment itself caused a 10- to 20-fold increase in the expression of both TAZ mRNA and protein over the ensuing 2 days in these cells, an effect that was blunted, but not eliminated, by siRNA treatment (Fig. 1C).

Thus, BMP-2 appears to stimulate a transcriptional program in which increased expression of TAZ, as well as subsequent TAZ-stimulated expression of the osteocalcin gene by Runx2, may be important for the osteoblast differentiation program. To test this idea, we performed immunoprecipitation experiments and verified a direct interaction of TAZ and Runx2 in transfected cells (Fig. 1D). TAZ expression caused a dose-dependent increase in Runx2-driven expression of an osteocalcin gene promoter fragment construct monitored in a luciferase reporter assay (Fig. 1E and fig. S2). 14-3-3 binds to TAZ through a site involving Ser^{89}, sequestering TAZ in the cytoplasm and limiting its ability to influence gene expression (*18*). Replacement of Ser^{89} with Ala (S89A mutation) disrupted 14-3-3 binding and relocalized TAZ preferentially within the nucleus, often in punctate foci (Fig. 1F) (*18*). When we used an S89A mutant of TAZ, the TAZ-dependent transcriptional coactivation of Runx2-driven gene expression was enhanced 2 to 3 times (Fig. 1E). Similar results for Runx2 transcriptional coactivation were obtained in C3H10T1/2 mouse embryonic fibroblasts (fig. S3). Both the interaction of TAZ with Runx2 and the stimulatory effect of TAZ on osteocalcin gene promoter activity were dependent on the TAZ WW domain (fig. S7). We also generated stable C2C12 cell lines expressing Flag-tagged TAZ and observed a BMP-2–dependent targeting of TAZ to the endogenous osteocalcin promoter by chromatin immunoprecipitation. (Fig. 1G). Taken together, these data indicate that TAZ is a 14-3-3–regulated transcriptional coactivator for

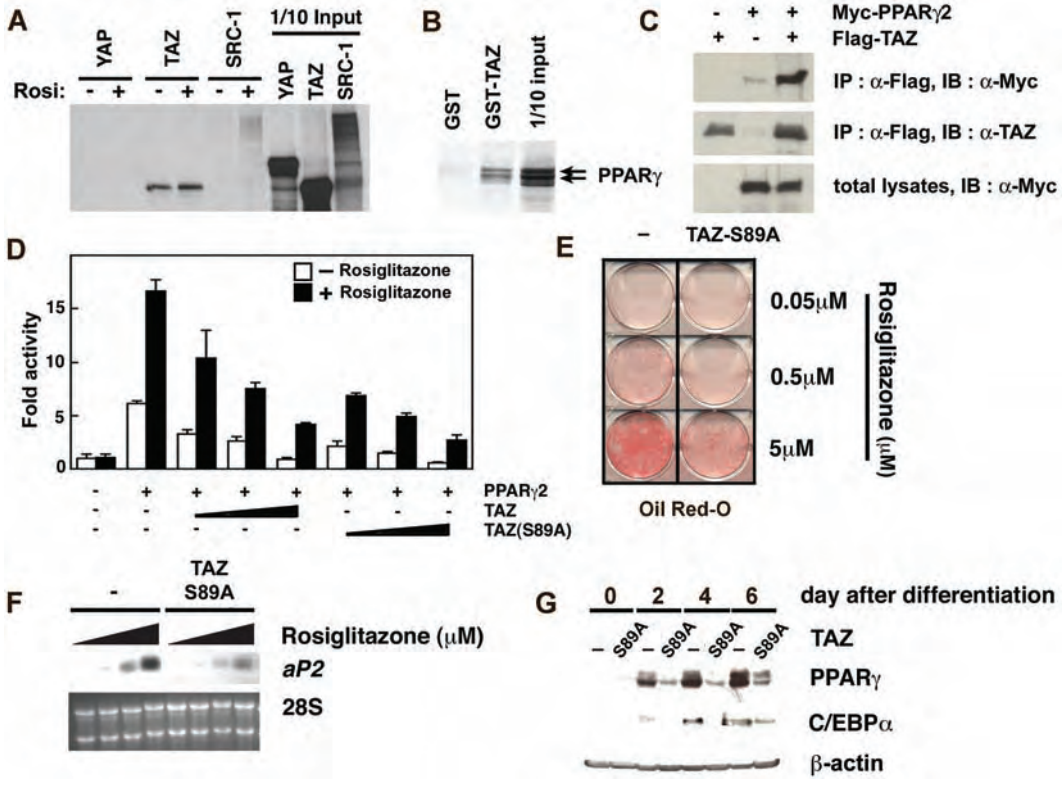


Fig. 2. TAZ binds to PPARγ and inhibits transcription from the aP2 promoter. (**A**) Binding of TAZ to glutathione *S*-transferase (GST) fused with PPARγ. Immobilized GST-PPARγ was incubated with in vitro translated, ^{35}S-labeled YAP or TAZ, or with the ligand-dependent PPARγ-binding protein SRC-1, in the presence or absence of 1 μM Rosiglitazone, and analyzed by autoradiography. (**B**) Binding of PPARγ to GST-TAZ. Immobilized GST-mTAZ was incubated with in vitro translated ^{35}S-labeled PPARγ. The PPARγ doublet represents sites of alternative translational initiation. (**C**) Interaction of TAZ with PPARγ in cells. 293T cells were transfected with Flag-tagged TAZ or Myc-tagged PPARγ and analyzed by coimmunoprecipitation. (**D**) Inhibition of PPARγ-driven gene expression by TAZ. U2OS cells were transfected with an aP2 promoter–containing firefly luciferase reporter construct (aP2-Luc), a PPARγ expression vector [SV-sport–PPARγ2, (*6*)] and various amounts of plasmids encoding either wt-TAZ or the S89A TAZ mutant (EF-mTAZ-NFLAG or EF-mTAZ(S89A)-NFLAG, respectively). Total DNA in each sample was equalized by addition of empty pEF-Bos vector DNA. Luciferase activity was measured after 24 hours in the presence or absence of 1 μM Rosiglitazone and normalized to the activity of a cotransfected *Renilla* luciferase reporter construct used as an internal control. (**E**) Inhibition of adipocyte differentiation by stable expression of TAZ. 3T3-L1 cells were infected with a control pBabe-puro retroviral vector (–) or a retroviral vector containing TAZ S89A. Puromycin-resistant cells were treated with Rosiglitazone and stained with Oil Red O and photographed at 8 days (*6*). (**F**) Expression of fatty acid–binding protein aP2. Total RNA was isolated from cells in (E), and probed with an adipocyte-specific cDNA to aP2. Ethidium bromide staining (bottom) verifies equal loading of RNA in each lane. (**G**) Inhibition of PPARγ-dependent gene expression by TAZ S89A. Retrovirally infected cells were lysed at the indicated times after treatment with 5 μM Rosiglitazone, and PPARγ and C/EBPα protein levels were analyzed by Western blotting.

Runx2-stimulated osteocalcin gene expression and an important endogenous regulator of osteoblast differentiation.

PPARγ, a member of the nuclear hormone receptor superfamily, is a ligand-activated transcription factor that is critical for adipocyte differentiation (*6*, *7*, *9*, *23*–*25*). Like Runx2, PPARγ also contains a Pro-Pro-X-Tyr WW domain–binding motif within its ligand-independent activation domain. We therefore investigated the functional interaction between TAZ and PPARγ. PPARγ bound in a ligand-independent manner to TAZ, but not to YAP, in experiments with tagged proteins in vitro and in transfected cells (Fig. 2A through C). In contrast to the stimulatory effect of TAZ on Runx2-driven gene expression, TAZ directly inhibited the ability of PPARγ to stimulate gene expression for the endogenous fatty acid–binding protein aP2, in both the presence and absence of the PPARγ-activating ligand Rosiglitazone, as assayed with a luciferase reporter construct (Fig. 2D). The inhibitory effect of TAZ on PPARγ-driven gene expression was even more pronounced when the constitutively nuclear S89A form of TAZ was used (Fig. 2D). Similar to what we observed with Runx2, the interaction of TAZ with PPARγ and the inhibitory effect of TAZ on aP2 gene expression required the TAZ WW domain (fig. S8).

To investigate the cellular role of TAZ on PPARγ-regulated genes, we generated stable 3T3-L1 cell lines expressing TAZ-S89A or a vector control, and we monitored adipocyte differentiation in response to Rosiglitazone. Enhanced expression of TAZ-S89A resulted in inhibition of adipocyte differentiation as revealed by suppression of Rosiglitazone-induced expression of aP2 and reduction in Oil Red O staining of the cells (Fig. 2, E and F). PPARγ participates in a positive feedback loop by driving transcription of its own gene, as well as that of other target genes such as the one for C/EBPα, as part of the normal adipocyte differentiation program (*26*). Overexpression of TAZ-S89A also inhibited these PPARγ-driven processes (Fig. 2G). Similar results were also observed with wild-type TAZ, although its effect was smaller. TAZ did not affect binding of PPARγ to its site on the aP2 promoter as seen from an oligonucleotide-binding assay (fig. S9). Instead, chromatin immunoprecipitation experiments localized TAZ to the endogenous aP2 promoter during these adipocyte differentiation experiments (Fig. 3D). Thus, TAZ may suppress adipocyte differentiation by transcriptionally repressing PPARγ-driven gene expression.

To further examine this possibility, we used siRNA to inhibit the production of endogenous TAZ in this same cell type (Fig. 3). After TAZ was depleted, the 3T3-L1 cells were placed in adipocyte differentiation medium containing various concentrations of Rosiglitazone. Cells were assayed 8 days later, by which

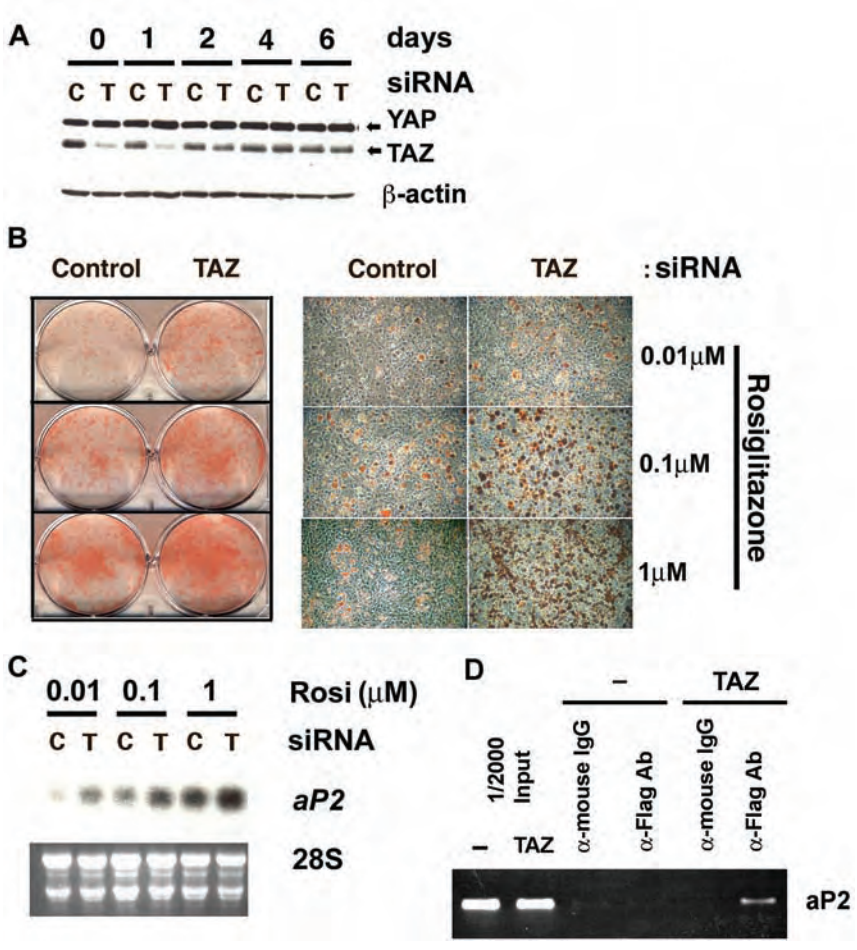
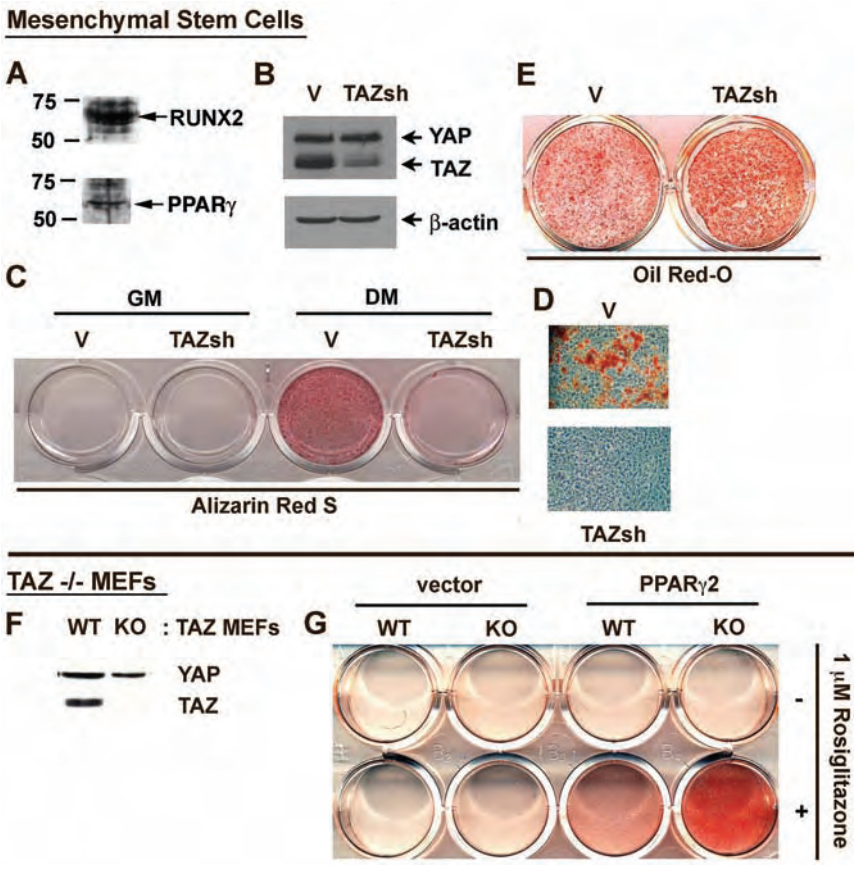


Fig. 3. Increased adipocyte differentiation from cells depleted of endogenous TAZ. (**A**) 3T3-L1 cells were transfected with control siRNA against GFP (C) or mTAZ-specific siRNA (T) for 36 hours and transferred to adipocyte differentiation medium. Cell lysates were examined for TAZ and YAP expression by Western blotting. (**B**) Adipogenesis in cells after inhibition of endogenous TAZ expression. As in (A), 3T3-L1 cells were treated with the indicated concentrations of Rosiglitazone for 8 days, stained with Oil Red O, and photographed. (**C**) Total RNA was isolated from cells shown in (B) and probed with an adipocyte-specific cDNA to aP2. (**D**) Stable 3T3-L1 cells expressing pBabe-puro retroviral vector control (–) or Flag-tagged wild-type TAZ were differentiated into adipocytes and analyzed by chromatin immunoprecipitation with the use of PCR primers for the endogenous aP2 promoter.

time TAZ expression was no longer inhibited. However, the transient reduction of expression of endogenous TAZ was sufficient to enhance ligand-induced adipocyte differentiation at each concentration of Rosiglitazone tested (Fig. 3, A through C). We conclude that TAZ is a transcriptional repressor of PPARγ-induced gene expression and an endogenous inhibitor of the adipocyte differentiation program.

Osteoblasts and adipocytes originate from the same MSCs through alternative activation of reciprocal transcriptional programs (*8*, *9*). Our findings suggested that TAZ may influence cell fate during the MSC differentiation process, because Runx2, PPARγ, and TAZ are present in these cells (Fig. 4, A and B; fig. S4). To test this, we isolated primary MSCs from murine bone marrow, allowed them to proliferate in culture, and infected them with control retroviruses or with retroviruses encoding short hairpin RNAs (shRNAs) against TAZ. Retrovirus-infected cells were selected after 7 days of culture in medium containing puromycin, and the entire surviving population of cells was assayed. Using this approach, we observed an ~75% reduction in TAZ protein expression (Fig. 4B). When these cells were subsequently cultured under conditions favoring osteoblast differentiation, the cells infected with the control retrovirus showed calcium deposition by 5 days (Fig. 4, C and D). In contrast, cells infected with the retroviruses encoding shRNAs against TAZ demonstrated almost no calcium deposition at this time, which revealed impairment of the osteoblast differentiation program. Conversely, when retrovirus-infected primary MSCs were cultured under conditions favoring adipocyte differentiation, the cells containing shRNA against TAZ demonstrated enhanced Oil Red O staining (Fig. 4E).

We wondered whether the increased adipogenic phenotype observed in cells lacking TAZ was blunted by the waning effectiveness and limited suppression of TAZ resulting from expression of siRNA or shRNA, respectively. We



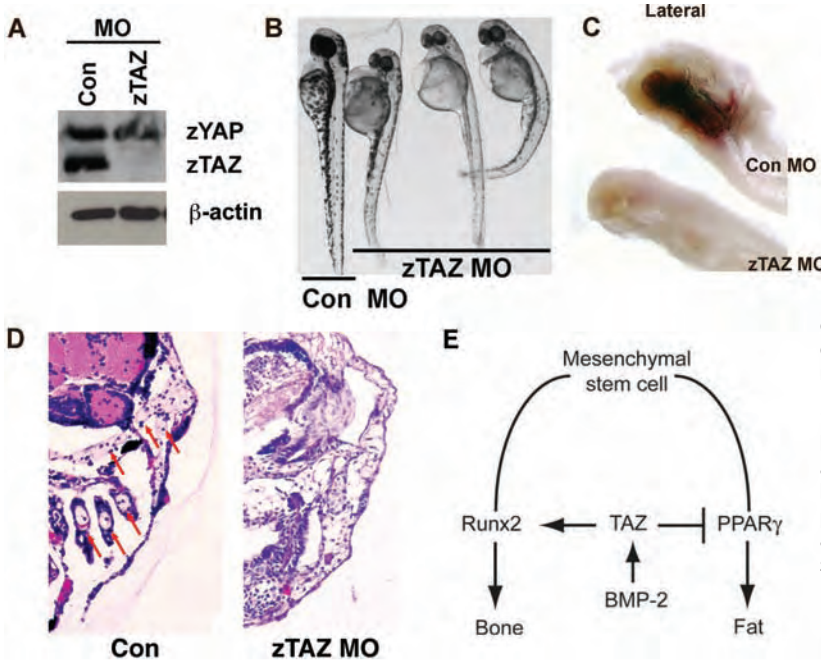


Fig. 5. Lack of skeletal ossification in TAZ-depleted zebrafish. (**A**) TAZ depletion in zebrafish embryos. Embryos were injected with a control morpholino oligomer (MO) or a TAZ-specific oligomer at the one- to two-cell stage and analyzed 1 to 2 days later by immunoblotting. (**B**) Morphology of TAZ-depleted zebrafish embryos at 2 dpf showing ventral curvature and pericardial edema. (**C**) Alizarin Red S staining of whole zebrafish embryos reveals absent skeletal ossification in TAZ-depleted embryos. Lateral and dorsal views of control- and TAZ-specific MO-injected embryos at 8 dpf. op: opercle; ch/bsr: ceratohyal and branchiostegal rays; cb5: ceratobranchial. (**D**) Hematoxylin-and-eosin staining of zebrafish sections reveals impaired bone development. Red arrows indicate normal bone morphology in control embryos, which is absent in TAZ-depleted animals. (**E**) A model for TAZ modulation of mesenchymal stem cell differentiation.

Fig. 4. Bone marrow-derived MSCs depleted of TAZ show decreased osteogenesis and increased adipogenesis. (**A**) Concurrent expression of Runx2 and PPARγ in primary bone marrow–derived MSCs. Nuclear extracts from MSCs were analyzed by Western blotting with the indicated antibodies. (**B**) Depletion of TAZ with shRNA. MSCs were infected with a pSRP retrovirus control (V) or pSRP encoding TAZ short hairpin RNA (TAZsh). After puromycin selection for 7 days, the MSC population was analyzed for depletion of endogenous TAZ by Western blotting. (**C** and **D**) Impaired osteoblast differentiation TAZ-depleted MSCs. After 7 days of selection in puromycin, amplified cells were transferred to medium containing 0.1 μM dexamethasone, 50 μg/ml ascorbic acid, and 10 mM β-glycerophosphate for 5 days to induce osteoblast differentiation. Plates were stained with Alizarin Red S to visualize intracellular calcium deposition and photographed. (**E**) Enhanced adipocyte differentiation in TAZ-depleted MSCs. MSCs were induced to differentiate into adipocytes by culture in medium containing 1 μM dexamethasone, 5 μg/ml insulin, and 1 μM Rosiglitazone for 8 days, stained with Oil Red O, and photographed. (**F**) TAZ expression in knockout MEFs. TAZ expression was analyzed in whole-cell lysates from wild-type and $TAZ^{-/-}$ MEFs by using an antibody that recognizes both TAZ and YAP. (**G**) Enhanced adipocyte differentiation of TAZ knockout MEFs. Wild-type and $TAZ^{-/-}$ MEFs were infected with a PPARγ2-expressing retrovirus or a retrovirus vector control, and the infected cells were selected by using puromycin. Cells were induced with differentiation medium in the presence or absence of 1 μM Rosiglitazone as above, stained with Oil Red O, and photographed.

therefore addressed adipocyte differentiation in wild-type mouse embryo fibroblasts (MEFs) and $TAZ^{-/-}$ MEFs generated by homologous recombination (*27*) (Fig. 4F). MEFs were immortalized, infected with a retrovirus expressing PPARγ2, and cultured in medium containing Rosiglitazone to induce adipogenesis. Adipogenesis was more pronounced in $TAZ^{-/-}$ MEFs than in wild-type cells (Fig. 4G and fig. S5), consistent with a negative role of endogenous TAZ in the adipocyte differentiation process.

To investigate the in vivo role of TAZ on bone development, we used a zebrafish vertebrate model system. We cloned the zebrafish TAZ ortholog (fig. S6) and used antisense morpholino oligomers injected at the one- to two-cell stage to decrease expression of TAZ (Fig. 5A). The TAZ-depleted embryos survived for up to 8 days after fertilization (dpf), and had developmental abnormalities including a ventral axis curvature and pericardial edema

(*28*) (Fig. 5B). Bone development was visualized in whole embryos by Alizarin Red S staining (Fig. 5C), along with hematoxylin-and-eosin staining of thin sections (Fig. 5D). In control animals, extensive skeletal development was evident in the cranial and pharyngeal region at 8 dpf. No bone formation was observed in any of the TAZ-depleted embryos at 8 dpf, the latest time point we could observe before embryonic death. These findings confirm a critical role for TAZ in osteoblast differentiation in vivo. We are unable to comment on the role of TAZ in adipogenesis in these embryos, because adipocytes have not been described in teleosts (although they presumably exist), and in other vertebrate species, fat deposition does not occur until the postnatal period (*23*).

One function of TAZ is as a transcriptional modifier of mesenchymal stem cell differentiation by promoting osteoblast differentiation while simultaneously impairing adipocyte differentiation, as we have shown (Fig. 5E). Differentiation of MSCs into osteoblasts is critically dependent on Runx2 (*4*, *5*). Our findings implicate TAZ in this process: (i) TAZ functions as an endogenous coactivator of Runx2 in cells; (ii) stimuli that promote bone formation transcriptionally up-regulate TAZ concurrently with Runx2; and (iii) TAZ-deficient zebrafish embryos are defective in bone formation. In contrast, stem cell differentiation into adipocytes requires PPARγ-dependent transcriptional events that are directly inhibited by endogenous TAZ. Thus, TAZ may act as a molecular rheostat to fine-tune the balance between osteoblast and adipocyte development.

References and Notes

1. M. F. Pittenger *et al.*, *Science* **284**, 143 (1999).
2. A. I. Caplan, S. P. Bruder, *Trends Mol. Med.* **7**, 259 (2001).
3. P. Ducy, R. Zhang, V. Geoffroy, A. L. Ridall, G. Karsenty, *Cell* **89**, 747 (1997).
4. T. Komori *et al.*, *Cell* **89**, 755 (1997).
5. F. Otto *et al.*, *Cell* **89**, 765 (1997).
6. P. Tontonoz, E. Hu, B. M. Spiegelman, *Cell* **79**, 1147 (1994).
7. E. D. Rosen *et al.*, *Mol. Cell* **4**, 611 (1999).
8. K. Nakashima, B. de Crombrugghe, *Trends Genet.* **19**, 458 (2003).
9. E. D. Rosen, C. J. Walkey, P. Puigserver, B. M. Spiegelman, *Genes Dev.* **14**, 1293 (2000).
10. P. Meunier, J. Aaron, C. Edouard, G. Vignon, *Clin. Orthop.* **80**, 147 (1971).
11. R. Burkhardt *et al.*, *Bone* **8**, 157 (1987).
12. O. Kajkenova *et al.*, *J. Bone Miner. Res.* **12**, 1772 (1997).
13. S. Verma, J. H. Rajaratnam, J. Denton, J. A. Hoyland, R. J. Byers, *J. Clin. Pathol.* **55**, 693 (2002).
14. M. E. Nuttall, J. M. Gimble, *Bone* **27**, 177 (2000).
15. F. S. Kaplan, E. M. Shore, *J. Bone Miner. Res.* **15**, 2084 (2000).
16. D. J. Prockop, *Science* **276**, 71 (1997).
17. S. C. Chang *et al.*, *J. Surg. Res.* **119**, 85 (2004).
18. F. Kanai *et al.*, *EMBO J.* **19**, 6778 (2000).
19. C. B. Cui, L. F. Cooper, X. Yang, G. Karsenty, I. Aukhil, *Mol. Cell. Biol.* **23**, 1004 (2003).
20. R. Yagi, L. F. Chen, K. Shigesada, Y. Murakami, Y. Ito, *EMBO J.* **18**, 2551 (1999).
21. T. Katagiri *et al.*, *J. Cell Biol.* **127**, 1755 (1994).
22. S. Gallea *et al.*, *Bone* **28**, 491 (2001).
23. P. Cornelius, O. A. MacDougald, M. D. Lane, *Annu. Rev. Nutr.* **14**, 99 (1994).
24. Y. Barak *et al.*, *Mol. Cell* **4**, 585 (1999).
25. N. Kubota *et al.*, *Mol. Cell* **4**, 597 (1999).
26. Z. Wu *et al.*, *Mol. Cell* **3**, 151 (1999).
27. Y. Tian, T. Benjamin, unpublished observations.
28. J.-H. Hong, N. Hopkins, M.B. Yaffe, unpublished observations.

29. Assistance from H. Jo and R. Nissen in zebrafish studies and S. Bissonnette in MSC studies is gratefully acknowledged. This work was supported by NIH grants CA042063 to P.A.S., and GM60594 and GM68762 and a Burroughs-Wellcome Career Development Award to M.B.Y.

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5737/1074/DC1
Materials and Methods
Figs. S1 to S9
References and Notes

11 February 2005; accepted 28 June 2005
10.1126/science.1110955

Formation of Regulatory Patterns During Signal Propagation in a Mammalian Cellular Network

Avi Ma'ayan,[1] Sherry L. Jenkins,[1] Susana Neves,[1] Anthony Hasseldine,[1] Elizabeth Grace,[1] Benjamin Dubin-Thaler,[3] Narat J. Eungdamrong,[1] Gehzi Weng,[1]* Prahlad T. Ram,[1]† J. Jeremy Rice,[4] Aaron Kershenbaum,[4] Gustavo A. Stolovitzky,[4] Robert D. Blitzer,[1,2] Ravi Iyengar[1]‡

We developed a model of 545 components (nodes) and 1259 interactions representing signaling pathways and cellular machines in the hippocampal CA1 neuron. Using graph theory methods, we analyzed ligand-induced signal flow through the system. Specification of input and output nodes allowed us to identify functional modules. Networking resulted in the emergence of regulatory motifs, such as positive and negative feedback and feedforward loops, that process information. Key regulators of plasticity were highly connected nodes required for the formation of regulatory motifs, indicating the potential importance of such motifs in determining cellular choices between homeostasis and plasticity.

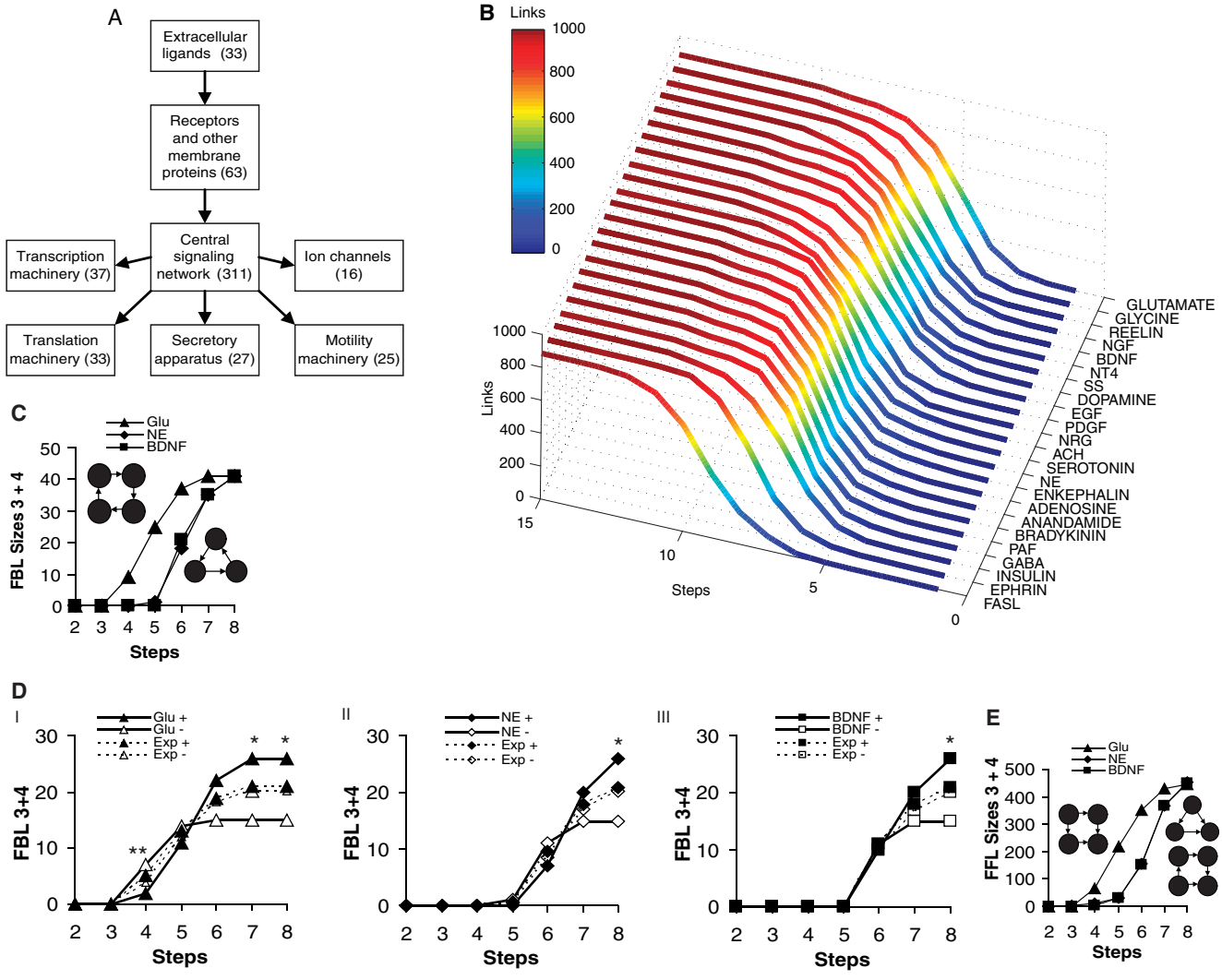
A mammalian cell may be considered as a central signaling network connected to various cellular machines that are responsible for phenotypic functions (*1*). Cellular machines such as transcriptional, translational, motility, and secretory machinery can be represented as sets of interacting components that form functional local networks. The central signaling network that connects the various machine networks also receives and processes signals from extracellular entities such as hormones or neurotransmitters and ions. Experimental work has defined how different pathways interact to form networks and small-scale regulatory configurations such as switches (*2*, *3*), gates (*4*, *5*), feedback loops (*6*, *7*), and feedforward motifs (*8*, *9*) that decode signal duration and strength and process information. Identifying and characterizing regulatory motifs can move us from thinking about individual components to considering the functions of groups of components that act in a coordinated manner. Understanding how the functional organization of cellular systems changes in response to information flow is an important goal in systems biology. For systems containing many components, obtaining an overview of the patterns of regulatory motifs and defining their interrelationships can provide a format for in-depth analysis of individual units using quantitative biochemical representations.

From data in the experimental literature, we constructed a system of interacting cellular components involved in phenotypic behavior and used graph theory methods (*10*–*12*) to analyze qualitative relationships between nodes (components) in a network. In signaling networks, activation is achieved as a response to a stimulus. Information propagates through the system by a series of coupled biochemical reactions to regulate components responsible for cellular phenotypic functions. Here, we identify the regulatory features that emerge during such information flow in a simplified representation of a mammalian hippocampal CA1 neuron. Such neurons are capable of plasticity as defined by their ability to undergo long-term potentiation of synaptic responses (*13*, *14*).

We represented the CA1 neuron as a set of interacting components that make up a network of signaling pathways that connects to various

Departments of [1]Pharmacology and Biological Chemistry and [2]Psychiatry, Mount Sinai School of Medicine, New York, NY 10029, USA. [3]Department of Biological Sciences, Columbia University, New York, NY 10029, USA. [4]Functional Genomics and Systems Biology, IBM T. J. Watson Research Center, Yorktown Heights, NY 10598, USA.

*Present address: Scios Inc., 6500 Paseo Padre Parkway, Fremont, CA 94555, USA.
†Present address: Department of Molecular Therapeutics, M. D. Anderson Cancer Center, Houston, TX 77025, USA.
‡To whom correspondence should be addressed. E-mail: Ravi.Iyengar@mssm.edu



	Motifs counts		
Motif #	CN*	SN**	Z-score
FBL3	22	8.6 ± 8.1	4.34
SCAF	25	9.0 ± 3.4	4.72
BIFAN	1009	180.3 ± 28.0	29.60
FFL4	301	103.4 ± 17.2	11.51

* CN- Cellular Network.
** SN- Shuffled Networks.
Mean ± SD computed for 100 shuffled networks.

Fig. 1. Assembly of a function-based cellular network. (**A**) Block diagram of information flow from ligands to cellular machines. Nodes are associated with functional compartments. The counts of nodes are in parentheses. (**B**) Total links in subnetworks created in steps from ligands. The total number of links accumulated as a signal moves through the steps, downstream from various ligands, is compared. (**C**) Total number of three- and four-component feedback loops (positive and negative) in subnetworks with two to eight steps from Glu, NE, and BDNF. (**D**) Counts of positive and negative feedback loops in subnetworks from Glu (I), NE (II), and BDNF (III). The counts in the biological subnetworks (solid lines) are compared with the expected counts (dashed lines) developed using combinatorial probabilities for positive or negative feedback loops based on the number of total positive and negative links. **, $P < .05$; *, $P < .08$; binomial test. (**E**) Counts of three- and four-component feedforward motifs in subnetworks with two to eight steps from Glu, NE, and BDNF. (**F**) An abbreviated list of the motifs found in the fully connected network using the MFinder program (*19*). CN, cellular network; SN, shuffled networks; SNs were used as controls. Under SN, the mean and SD are given for motifs in 100 shuffled networks. See (*15*) for details.

cellular machines (Fig. 1A and fig. S1) (*15*). A fully connected network was constructed from direct interactions with functional effects documented in the experimental literature (fig. S2). We identified various regulatory motifs in the network. Some motifs were statistically enriched (Z test) compared with motifs found in shuffled networks with similar connectivity distribution (fig. S3 and table S1). We studied signal propagation resulting from ligand occupancy of receptors by building and analyzing a series of subnetworks that originate from nodes representing ligands. This is termed pseudodyamics because it represents propagation of reactions in chemical space rather than time series. Direct interaction between any two components, termed a link, consists of one or more underlying chemical reactions. Signal propagation from node to node is organized within the chemical space in units of steps. For any given node at step n, all immediate upstream nodes are at step n–1 and all nodes positioned one link downstream are at step n+1.

We counted the number of links per step as signals propagate from ligand-receptor interactions to their downstream effectors. The analysis of the emergent subnetworks for the different ligands showed a discernible pattern. For ligands that cause rapid, transient changes, such as glutamate and glycine, which regulate the Ca^{2+}-permeable *N*-methyl-D-aspartate (NMDA)–type glutamate receptor, early signal branching was

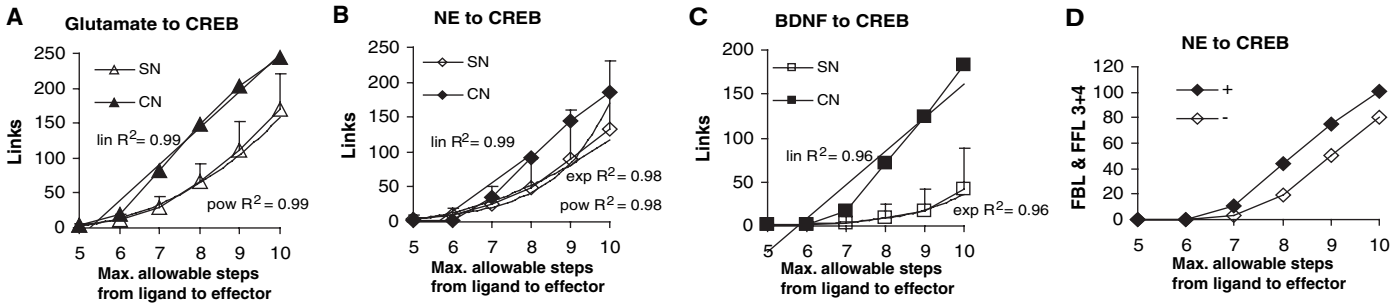


Fig. 2. Characteristics of subnetworks from three ligands to CREB. Subnetworks were analyzed starting from three extracellular ligands—glutamate, NE, and BDNF (source nodes)—extending to the transcription factor CREB (target node). (**A** to **C**) Changes in the number of links with increasing number of steps to reach the effectors. The same analysis was done with shuffled networks. Only the directionality of the links that do not involve the ligands or the effectors was randomly swapped while preserving the connectivity. The resultant graphs for both the cellular network (CN) and the shuffled networks (SN) were curve-fitted with Microsoft Excel. For all of the cellular subnetworks, the best fit function was linear (lin), R^2 = 0.96 to 0.99. For all of the shuffled networks, the best fit was obtained with exponential or power-law functions (exp or pow), R^2 = 0.96 to 0.99. (**D**) Counts of total three- and four-node positive and negative feedback and feedforward loops in subnetworks when different maximum numbers of steps are specified between NE and CREB.

extensive (that is, many links were formed in relatively few steps) (Fig. 1B). For ligands that cause permanent changes, such as FasL, which induces apoptosis, or ephrin, which alters neuronal morphology, there were fewer branches as the signal moved through the network. Between these extremes were many growth factors and ligands that bind G protein–coupled receptors (Fig. 1B). When the signal originating from any ligand had progressed through 15 steps, most of the network (nearly 1000 links) was engaged. This is a common property of large, highly connected directed graphs and in our case represents a very extensive propagation of the signal from the ligand analogous to prolonged receptor activation. For any individual ligand, the whole network was never fully affected, with a few nodes (usually other ligands) with single directed outgoing interactions not engaged. We characterized the network that emerged as signals were traced through each step from three ligands that are key regulators of plasticity in hippocampal neurons: glutamate (*16*), norepinephrine (NE) (*17*), and brain-derived neurotrophic factor (BDNF) (*18*). Glutamate influenced more links and nodes in early stages than did NE or BDNF, which showed similar profiles (fig. S4, A and B). The subnetwork characteristics were similar (fig. S4, C to E).

We analyzed the types of regulatory motifs formed as signals propagate from the ligands. A motif is a group of interacting components capable of signal processing. Motifs such as positive feedback loops promote the persistence of signals and serve as information storage devices (*2*), whereas negative loops limit signal propagation through the network. In counting these motifs, we used all possible configurations of loops with three and four components (Fig. 1C). In our model, glutamate activates 25 feedback loops within five steps, whereas NE and BDNF require six steps to recruit 20 feedback loops (Fig. 1C). We determined the relative abundances of negative and positive feedback loops as signal propagated from glutamate, NE, or BDNF (Fig. 1D). Using a bionomial test and the combinatorial possibilities arising from the ratios of positive to negative links, it appeared that there were more than the expected number of negative feedback loops compared with positive feedback loops for glutamate at early steps ($P < 0.05$). This indicates that the early steps may have built-in controls to limit signal propagation. As the signal progressed to later steps, positive feedback loops tended to be more abundant than expected for all three ligands ($P < 0.08$). These profiles suggest that weak or short-lived signals may not penetrate into the network because of the early barrier posed by the abundance of negative feedback loops. However, as the number of steps increases and positive feedback loops appear to be more abundant, signals should persist and be able to evoke a biological response.

We also counted feedforward motifs (Fig. 1E and fig. S5), scaffolds (fig. S6A-I), and bifans (fig. S7B-I). Scaffold motifs are sets of three nodes connected through three neutral links. Their presence may indicate a mechanism for local clustering and may represent the basis for spatial specification of information flow. Feedforward loops are three- or four-component motifs in which the signal splits at a source node and consolidates at a target node. Feedforward loops can be either positive or negative (*15*). Positive feedforward loops provide two potential functions: a redundant set of pathways for information flow and an ability to extend the duration of activation or inhibition of the affected downstream component. When feedforward loops contain inhibitory links, they may function as gates (*2*). Bifans are four-component motifs in which two upstream components coregulate two downstream components (Fig. 1F) (*19*).

For each of these motifs, glutamate required fewer steps than did NE or BDNF to engage more motifs, although by step 8, when maximal connectivity is obtained, signals from the three ligands had spread to the same number of motifs (Fig. 1E and figs. S5 and S6). There was an increasing preponderance of positive over negative feedforward motifs for glutamate, NE, and BDNF as the signal propagated through eight steps (fig. S5). The overall profile for any feedforward loops (positive or negative) was similar for all three ligands, although glutamate reached more feedforward loops in fewer steps. The bifan motifs were as abundant as feedforward loops and showed a similar pattern (fig. S6C).

In the CA1 neuron, signals from receptors affect major effectors such as the AMPA receptor channel (AMPAR), which regulates excitatory postsynaptic potentials, and the transcription factor cyclic adenosine monophosphate (cAMP) response element–binding protein (CREB), which regulates transcription. We analyzed a series of subnetworks that extend from glutamate, NE, and BDNF to these two effectors (figs. S7A and S8A) by varying the number of steps needed to reach the effectors from the ligand. The number of nodes engaged per step was nearly linear for all three ligands (Fig. 2, A to C). In contrast, when such pathways were tracked in subnetworks with randomly shuffled connectivity (*15*), the increase in links per step was best fit by either an exponential or a power-law function, suggesting that the binary interactions within the cellular system may be specified to provide preferential routes to key effectors that define the phenotype of the cell. Analysis of these subnetworks (*15*) indicated that even the most highly connected nodes only used some of their links to function within the preferred paths. Looking for the shortest paths from an input node (a ligand such as NE) to an output node (an effector such as CREB) allowed us to identify functional modules defined here as a set of interactions that may carry out a specific function and act semi-independently from other functional modules. Further refinement by inclusion of spatial specification and temporal dynamics is needed to define such modules rigorously.

The counts of regulatory motifs formed within the subnetworks, as signal propagates were nearly linear for both the positive and negative motifs. For glutamate and BDNF in the subnetworks leading to CREB, the positive

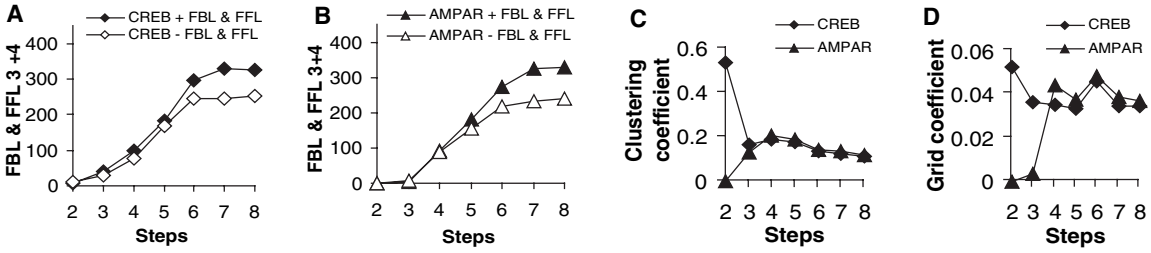


Fig. 3. Comparison of motifs and clustering in subnetworks upstream from two effectors, CREB and AMPAR. (**A** and **B**) Positive and negative feedback loops (FBL) and feedforward loops (FFL) in subnetworks as the network is extended by including more steps upstream of the effectors CREB or AMPAR. (**C** and **D**) Changes in clustering coefficient and grid coefficient in these subnetworks.

and negative motifs were evenly balanced through nine steps (fig. S8, A and B). Such balance of regulatory loops might explain how homeostasis is achieved even when a perturbing signal propagates through the system. Positive feedback and feedforward loops were more abundant than negative loops in the subnetworks from NE to CREB (Fig. 2D). This may allow the subnetworks to persistently hold and transfer information causing long-term changes. Such configurations of motifs may provide a systems-level explanation of why the cAMP pathway is associated with the late phase of long-term potentiation (LTP) in the CA1 neuron (*20*–*22*). Similar patterns of positive to negative motifs were observed for signal propagation from NE to AMPAR (fig. S9F). Patterns from glutamate and BDNF to AMPAR were also similar to that seen in subnetworks to CREB (fig. S9, E and G).

To analyze the subnetworks upstream of the two key effectors CREB and AMPAR, we used these effectors as source nodes and expanded the network in a stepwise manner, tracing the signals that feed into CREB and AMPAR. The number of positive and negative feedback loops converging on CREB and AMPAR were equal up to four steps upstream of the effector (Fig. 3, A and B). However, we found distinct patterns of clustering. Two steps above CREB, a high (0.53) clustering coefficient was observed. Clustering coefficients measure the abundance of three component motifs and thus indicate local connectivity density. By three or four steps upstream, for both CREB and AMPAR, the clustering coefficients were both above the average observed clustering for the entire network (0.2 versus 0.11) (Fig. 3C). The grid coefficients [extensions of the clustering coefficient that also consider rectangles (*23*)] for the subnetworks two to four steps upstream of CREB and AMPAR were also higher than that observed for the network as a whole (0.053 to 0.031 versus 0.026) (Fig. 3D). Extensive local communication between nodes upstream of key effectors may provide homeostatic regulation of these effectors.

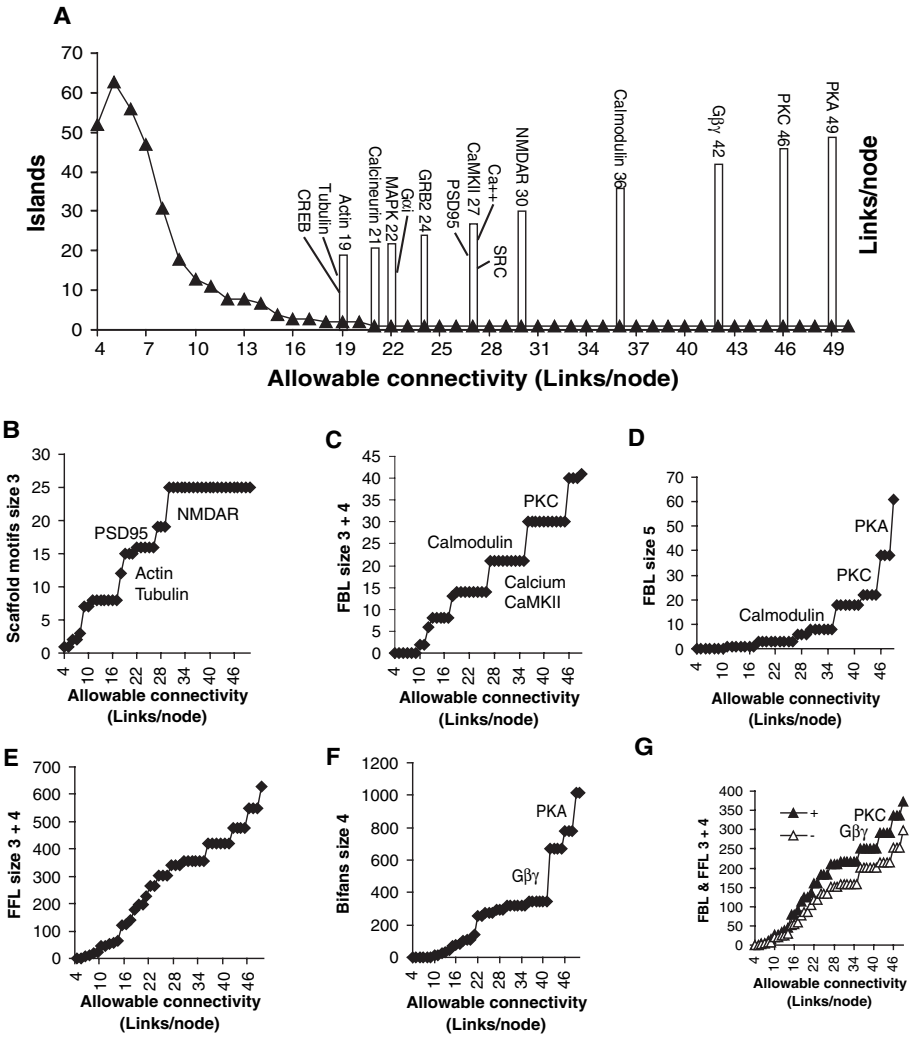


Fig. 4. Analysis of subnetworks with sequential incorporation of nodes with increasing connectivity. (**A**) Islands (isolated clusters of connected nodes) are plotted as a function of connectivity per node. A single network (one island) is formed when the nodes with 21 or fewer connections are included. All highly connected nodes starting at 19 links per node are shown. Single isolated nodes were not considered islands. (**B** to **F**) The number of motifs in the subnetworks created with nodes with different connectivity. Nodes that contribute disproportionately to the number of added motifs are listed next to the step they produce in the graph. (**G**) Counts of three- and four-component positive or negative feedback and feedforward loops are compared.

To evaluate the role of the highly connected nodes, we analyzed a series of subnetworks generated by the progressive inclusion of nodes with higher connectivities. We started with a sparse network in which we included only nodes with four or fewer links. We measured the number of islands (isolated clusters of nodes that lack paths between them), clustering, characteristic path length, and number of motifs as we gradually added nodes with higher connectivity (fig. S10, A and B). The system was initially highly fragmented (63 islands). When all of the nodes with up to 21 connections were included, the network became a single island (one network) (Fig. 4A). At this point, the most highly connected nodes, many of which are crucial components for LTP in hippocampal neurons, were not included. The nodes with more than 21 links per node included the four major protein kinases: mitogen-activated protein kinase (MAPK), calcium-calmodulin dependent protein

kinase II (CaMKII), protein kinase A (PKA), and protein kinase C (PKC) (Fig. 4A). Such highly connected nodes might contribute to regulatory motifs. Hence, we identified motifs that were formed if the highly connected nodes were included. The contributions of specific nodes to the formation of different motifs were readily discernable. Scaffolding motifs were largely dependent on the cytoskeletal proteins actin and tubulin, the synaptic scaffolding protein PSD-95, and NMDA receptors (Fig. 4B). Nearly 65% of the scaffolding motifs were formed without including enough nodes for the system to coalesce into a single island. In contrast, just 35% of the feedback and feedforward motifs and 20% of the bifan motifs were formed only in a network that included nodes with connectivity up to 21 links per node (Fig. 4, C to F). PKA and PKC appear to contribute to nearly 60% of the five-component feedback loops. Highly connected nodes, such as PKC, favored the emergence of positive motifs (Fig. 4D). These observations suggest that a function of these highly connected nodes may be to promote the formation of regulatory motifs that can allow for persistence of information and thus facilitate state change when appropriate external signals are received.

We developed maps to represent the pseudodynamic regulatory topology that our analyses had identified. We plotted the density of motifs at each step for the three ligands glutamate, NE, and BDNF by defining a "density of information processing" (DIP), which is the ratio of the additional new motifs (feedback loops, feedforward loops, and bifans) formed in each step to the increase in new links in each step (*15*) multiplied by the grid coefficient. This measure allowed us to identify the intensity and position in chemical space of information-processing activities. The DIP profile (Fig. 5A) at each step is plotted for the three different ligands through eight steps as signal propagates from receptors to cellular machines. Each DIP profile is distinctive, indicating that these represent differential stimulation of partially overlapping regions of the cellular network. Each ligand showed a "hot zone," where extensive information processing may occur. We also developed detailed maps to determine the position of the regulatory motifs in the chemical space of the network. For this, we specified the location of the nodes that participate in a motif between extracellular ligands and cellular machines on the basis of the shortest path lengths from the node within the motif to all extracellular ligands and to all components in the specified cellular machine. For each node, this yielded a "node location index" that measures the functional distance to each of the five cellular machines and to the ligands. We then identified the participation of these nodes in the various motifs. A parameter termed "motif location index" (MLI) was defined as the average of the location indices for the various nodes that comprise the motif in relation to the distance from the specified machine (*15*). Five maps corresponding to the different cellular machines were generated (Fig. 5B). These maps indicate the location of the various regulatory motifs between extracellular ligands and cellular machines. Both common and distinctive features are observed. When pathways from ligands to each of the cellular machines were considered, a higher density of regulatory motifs was found at the middle of the maps (note the red band in each map), indicating that a major portion of the information processing occurs at the "center" of the network.

Distinct patterns of motifs were observed upstream of the different cellular machines. Directly upstream of the transcriptional machinery, feedforward motifs were abundant. In contrast, for the translational machinery, the regulation was more distal. For the secretory machinery, feedforward and feedback loops and scaffolds were found to be closer to the machine. For both the motility machinery and ion channels, regulation was largely concentrated in the center of the network (around MLI + 0.5). Some regulatory motifs are made of components within the cellular machines themselves.

This study provides an initial view of the regulatory capabilities that are formed as information flows through a cellular network. The overall profile of emergent motifs can be an indicator of the cell's information-processing capability. Here, organization is defined solely in terms of chemical space. It will be necessary to integrate this description with the physical

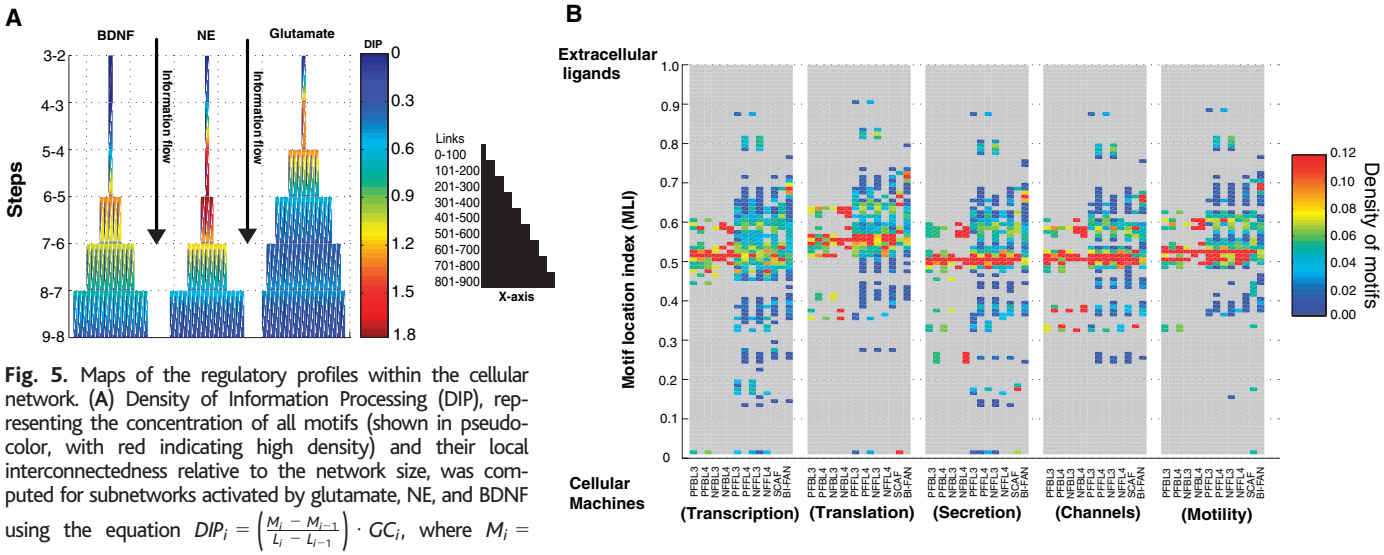


Fig. 5. Maps of the regulatory profiles within the cellular network. **(A)** Density of Information Processing (DIP), representing the concentration of all motifs (shown in pseudocolor, with red indicating high density) and their local interconnectedness relative to the network size, was computed for subnetworks activated by glutamate, NE, and BDNF using the equation $DIP_i = \left(\frac{M_i - M_{i-1}}{L_i - L_{i-1}}\right) \cdot GC_i$, where $M_i = FBL3_i + FBL4_i + FFL3_i + FFL4_i + BIFAN_i$. M_i is the total number of feedback loops, feedforward loops, and bifan motifs; L_i is the total number of links; and i is the step. *FBL3* and *FBL4* are feedback loops of size 3 and 4, *FFL3* and *FFL4* are feedforward loops of size 3 and 4, and *BIFAN* are bifan motifs of size 4. GC is the grid coefficient representing interconnectedness for the motifs, computed for the subnetwork at step *i*. DIP is plotted as a signal propagates vectorially through the network, as indicated by the downward arrow. The width of the bar represents the number of links engaged. **(B)** Relative distribution of regulatory motifs between ligands and cellular machines. Motifs were placed between extracellular ligands and cellular machines using the equation $MLI = \frac{\sum_{i=1}^{n}\left(\frac{CPLM_i}{CPLM_i + CPLL_i}\right)}{n}$ where n is the size of the motif, *CPLM* is the characteristic path length from a node within the motif to all other nodes in the cellular machine, and *CPLL* is the characteristic path length from a node to all extracellular ligands. If a node is an extracellular ligand, *CPLL* = 0 for that node; if the node is in the plasma membrane, *CPLL* = 1. If a node belongs to a cellular machine, *CPLM* = 0 for that node. The counts of motifs were placed in 100 bins based on the computed MLI and normalized to the fraction of total motifs for each class of motifs. An MLI value of 0 represents location at the cellular machines where all the nodes that make up the motif are within the cellular machine; a value of 1.0 represents location at the ligand level. PFBL and NFBL, three- and four-node positive and negative feedback loops, respectively; PFFL and NFFL, three- and four-node positive and negative feedforward loops; SCAF, three-node scaffold motifs; BIFAN, four-node bifan motifs. See (*15*) for details.

compartmentalization of components within the cell, and with temporal profiles of activation, to obtain a more comprehensive picture of the functional organization of a cell. Still, distance in chemical space (i.e., number of links between two distal nodes) is likely to be a major determinant of information processing that regulates phenotypic behavior.

The maps for individual ligands or cellular machines show distinct patterns of motifs. Combinations of ligands will likely produce many more patterns of connectivity. Thus, a cellular system may not be a single network but rather an ensemble of network configurations that are evoked by the stimuli-induced activation of various parts of the system. Identifying these network configurations and the functions they evoke is likely to provide more complete descriptions of how molecular interactions lead to cellular choices between homeostasis and plasticity.

References and Notes

1. J. D. Jordan, E. M. Landau, R. Iyengar, *Cell* **103**, 193 (2000).
2. U. S. Bhalla, R. Iyengar, *Science* **283**, 381 (1999).
3. U. S. Bhalla, P. T. Ram, R. Iyengar, *Science* **297**, 1018 (2002).
4. R. Iyengar, *Science* **271**, 461 (1996).
5. R. D. Blitzer *et al.*, *Science* **280**, 1940 (1998).
6. G. Lahav *et al.*, *Nat. Genet.* **36**, 147 (2004).
7. D. Angeli, J. E. Ferrell, E. D. Sontag, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1822 (2004).
8. S. Mangan, A. Zaslaver, U. Alon, *J. Mol. Biol.* **334**, 197 (2003).
9. S. Mangan, U. Alon, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11980 (2003).
10. D. J. Watts, S. H. Strogatz, *Nature* **393**, 440 (1998).
11. L. A. Amaral *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11149 (2000).
12. A. L. Barabasi, R. Albert, *Science* **286**, 509 (1999).
13. T. V. Bliss, G. L. Collingridge, *Nature* **361**, 31 (1993).
14. S. A. Siegelbaum, E. R. Kandel, *Curr. Opin. Neurobiol.* **1**, 113 (1991).
15. Materials and methods are available as supporting material on *Science* Online.
16. O. Hvalby *et al.*, *Experientia* **43**, 599 (1987).
17. H. Katsuki, Y. Izumi, C. F. Zorumski, *J. Neurophysiol.* **77**, 3013 (1997).
18. H. Kang, E. M. Schuman, *Science* **267**, 1658 (1995).
19. N. Kashtan, S. Itzkovitz, R. Milo, U. Alon, *Bioinformatics* **20**, 1746 (2004).
20. P. V. Nguyen, T. Abel, E. R. Kandel, *Science* **265**, 1104 (1994).
21. R. Bourtchuladze *et al.*, *Cell* **79**, 59 (1994).
22. C. Pittenger, Y. Y. Huang, R. F. Paletzki *et al.*, *Neuron* **34**, 447 (2002).
23. G. Caldarelli *et al.*, *European Physical Journal B.* **38**, 183 (2004).

24. This research is supported by GM-54508 and GM-072853 from NIH and by an Advanced Research Center grant from the New York State Office of Science and Technology. A.M. is supported by Pharmacological Sciences Training grant GM-62754. S.N. is the recipient of an individual predoctoral National Research Service Award (GM-65065). R.D.B. is supported by National Institute on Drug Abuse grant DA15863. We thank B. Obrink, H. Dohlman, N. Hao, and J. Lisman for comments on the manuscript, M. Diverse-Pierluissi for help in identifying components of the secretory machine, S. Purushothaman for implementing the grid-coefficient function, and G. Kossinets (D. Watts laboratory, Columbia University, NY) for help with the initial analysis. Author contributions are described in the Supporting Online Material.

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5737/1078/DC1
Materials and Methods
SOM Text
Figs. S1 to S11
Tables S1 to S3
References

20 December 2004; accepted 7 July 2005
10.1126/science.1108876

Containing Pandemic Influenza at the Source

Ira M. Longini Jr.,[1]* Azhar Nizam,[1] Shufu Xu,[1] Kumnuan Ungchusak,[2] Wanna Hanshaoworakul,[2] Derek A. T. Cummings,[3] M. Elizabeth Halloran[1]

Highly pathogenic avian influenza A (subtype H5N1) is threatening to cause a human pandemic of potentially devastating proportions. We used a stochastic influenza simulation model for rural Southeast Asia to investigate the effectiveness of targeted antiviral prophylaxis, quarantine, and pre-vaccination in containing an emerging influenza strain at the source. If the basic reproductive number (R_0) was below 1.60, our simulations showed that a prepared response with targeted antivirals would have a high probability of containing the disease. In that case, an antiviral agent stockpile on the order of 100,000 to 1 million courses for treatment and prophylaxis would be sufficient. If pre-vaccination occurred, then targeted antiviral prophylaxis could be effective for containing strains with an R_0 as high as 2.1. Combinations of targeted antiviral prophylaxis, pre-vaccination, and quarantine could contain strains with an R_0 as high as 2.4.

The world may be on the brink of an influenza pandemic (*1–4*). Avian influenza A (subtype H5N1) is causing widespread outbreaks among poultry in Southeast (SE) Asia, with sporadic transmission from birds to humans (*5*) and limited probable human-to-human transmission (*6*). Should an avian virus reassort with a human virus, such as influenza A subtype H3N2, within a dually infected human host or reassort in a nonhuman mammalian species, or if mutation of the virus occurs, the resulting new variant could be capable of sustained human-to-human transmission. The outbreak among humans would then spread worldwide via the global transportation network more rapidly than adequate supplies of vaccine matched to the new variant could be manufactured and distributed (*1*, *7*). The pressing public health questions are whether and how we can contain the spread of an emerging strain at the source or at least slow the initial spread to give time for vaccine development. We used a discrete-time stochastic simulation model of influenza spread within a structured geographically distributed population of 500,000 people in SE Asia to compare the effectiveness of various intervention strategies against a new strain of influenza. Here we examine the effectiveness of the targeted use of influenza antiviral agents (*8–12*), quarantine, and pre-vaccination with a poorly matched, low-efficacy vaccine in containing the spread of the disease at the source.

We used information about rural SE Asia (*13, 14*) to construct the model population. Our goal was to represent the contact connectivity of a typical rural SE Asian population. The model population of 500,000 people was distributed across a space of 5625 km^2, yielding a density of 89/km^2, which is approximately the population density of rural SE Asia (*13*). The 500,000 people were partitioned into 36 geographic localities. This model is an extension of a model used to simulate interventions against pandemic influenza in the United States (*12*).

The model [see the supporting online material (SOM) for details] represents the number of close and casual contacts that a typical person makes in the course of a day. The age and household size distributions of the population are based on the Thai 2000 census (*13*). Many of the mixing group sizes and distributions are based on a social network study of the Nang Rong District in rural Thailand (*14*). We constructed the social network for contacts sufficient to transmit influenza as a large set of connected mixing groups. The close contact groups consist of households, household clusters, preschool groups, schools, and workplaces; and the casual contact groups consist of other social settings (such as markets, shops, and temples) and a single regional 40-bed hospital. All people can

[1]Department of Biostatistics, The Rollins School of Public Health, Emory University, 1518 Clifton Road, N.E., Atlanta, GA 30322, USA. [2]Ministry of Public Health, Nonthaburi, Thailand. [3]Department of International Health, The Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA.

*To whom correspondence should be addressed. E-mail: longini@sph.emory.edu

mix in their households and within clusters of households, whereas children mix in preschool groups or schools according to their age and the probability that they are still in school. Children are assigned to schools across the geographic space according to the Nang Rong study. Adults mix in workplaces according to a distance function that distributes them across the geographic space informed by the Nang Rong study and national migration statistics (*15*, *16*). One concern for containment is that infected people might leave the modeled 500,000-person rural area. We estimate that the daily probability that a person will leave (escape) the area is on the order of 10^{-3} (*15*). The population structure and the resulting social network graphs and statistics are given in the SOM.

The natural history of influenza (Fig. 1A) has been relatively invariant over the past two pandemics and during the interpandemic period since 1968 (*12*, *17*–*20*). Calibration of the model requires information about both the relative and absolute magnitudes of the age-specific illness attack rates. Because of uncertainty about the relative age-specific illness attack rates for a future influenza pandemic in SE Asia, we calibrated the epidemic to a pattern that falls between two extremes. At one extreme, children would have a much higher illness attack rate than adults, the pattern observed during the 1957–1958 A (subtype H2N2) Asian influenza pandemic in the United States (*17*, *21*, *22*). At the other extreme, all age groups would have roughly the same illness attack rates, the pattern observed during the 1968–1969 A (H3N2) Hong Kong influenza pandemic in the United States (*17*, *22*–*24*). The pattern for interpandemic influenza in SE Asia appears to be more like the A (H2N2) pattern (*25*). We used the pattern shown in Fig. 1B.

The magnitude of the illness attack rates will depend on the unknown transmissibility of the new strain. The overall illness attack rate for the past Asian and Hong Kong pandemics was about 33% in the first wave. We calibrated the model with a target overall illness attack rate of 33%, corresponding to a basic reproductive number (R_0) (the average number of secondary infections caused by a single typical infectious individual in a completely susceptible population) of 1.4 (see the SOM). By varying the per-contact probability of infection in the model, we alter the R_0. Figure 1B shows the age-specific attack rates at R_0 values ranging from 1.1 to 2.4. For calibration to historical attack rates, influenza was introduced by randomly assigning 12 initial infectives. We simulated the emergence of a new influenza strain by introducing a single randomly assigned infective.

Intervention is triggered by the first case (that is, symptomatic infection), with a delay of 7, 14, or 21 days to implementation. This delay can be interpreted as a delay in recognition of illness, a delay in implementation of intervention, initiation of transmission by more than one initial infection, or a combination of these three factors. A sensitivity analysis considers delays up to 56 days (fig. S14). Once intervention begins, intervention in additional localities is implemented 1 day after the first case in the affected locality.

Targeted antiviral prophylaxis (TAP) is carried out by treating identified index cases (the first symptomatic illness in a mixing group) and offering prophylaxis only to the contacts of these index cases in predefined close contact groups (*12*); namely, households, neighborhood clusters, preschool groups, schools, and workplaces. Index cases are therapeutically treated the day after the onset of illness, and prophylaxis of contacts begins at the same time, both being given a single course of oseltamivir. A susceptible individual may receive subsequent courses if exposed to further index cases. We assume that a certain percent, varied in a sensitivity analysis, of household (preschool) index cases could be ascertained and that all their other household (preschool) members would receive prophylaxis (fig. S11). For index cases in a school or workplace, only

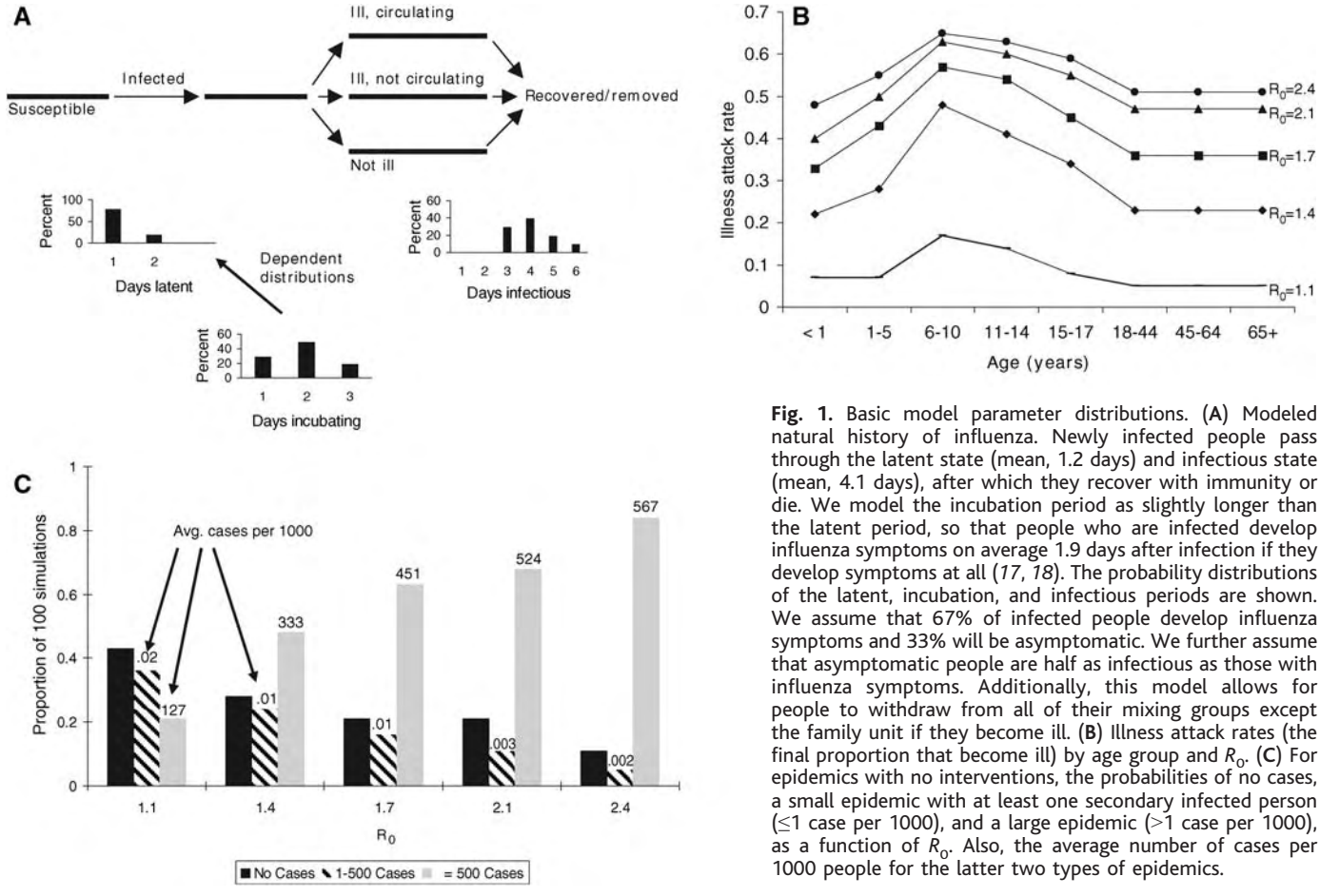


Fig. 1. Basic model parameter distributions. (**A**) Modeled natural history of influenza. Newly infected people pass through the latent state (mean, 1.2 days) and infectious state (mean, 4.1 days), after which they recover with immunity or die. We model the incubation period as slightly longer than the latent period, so that people who are infected develop influenza symptoms on average 1.9 days after infection if they develop symptoms at all (*17*, *18*). The probability distributions of the latent, incubation, and infectious periods are shown. We assume that 67% of infected people develop influenza symptoms and 33% will be asymptomatic. We further assume that asymptomatic people are half as infectious as those with influenza symptoms. Additionally, this model allows for people to withdraw from all of their mixing groups except the family unit if they become ill. (**B**) Illness attack rates (the final proportion that become ill) by age group and R_0. (**C**) For epidemics with no interventions, the probabilities of no cases, a small epidemic with at least one secondary infected person (≤1 case per 1000), and a large epidemic (>1 case per 1000), as a function of R_0. Also, the average number of cases per 1000 people for the latter two types of epidemics.

a certain percent of the people in that mixing group would receive prophylaxis. We used current estimates of the antiviral efficacy (AVE) of oseltamivir (*26*–*29*) (see the SOM).

The primary difficulty in TAP would be the identification of index cases. Because TAP is aimed at predefined close contact groups, the identification of potential TAP recipients would be less difficult than in classical contact tracing. An alternative and less resource-intensive strategy would be geographically targeted antiviral prophylaxis (GTAP), also known as ring prophylaxis. In this strategy, once an influenza case is identified in a locality, then a percentage, varied in a sensitivity analysis (fig. S12), of people in an entire locality are given one course of oseltamivir.

Household quarantine, like GTAP, is implemented within localities. The first case in a locality triggers a quarantine policy. Every case and a certain percentage of susceptible people restrict their movement to within their household and their neighborhood cluster. Because quarantined people would have more contact with their household and neighborhood contacts, the contact probabilities within households and household clusters are doubled for quarantined people.

A human influenza A (H5N1) vaccine is currently being tested (*7*) and may be available, but could be poorly matched to the emerging strain and thus be of low efficacy. For the model scenarios that use vaccination, we assume that pre-vaccination takes place long enough before the pandemic that vaccinated people can develop immunity. We assume a low vaccine efficacy for susceptibility (VE_S) (*30*) of 0.30 and a vaccine efficacy for infectiousness (VE_I) of 0.50. We carried out a sensitivity analysis on VE_I (fig. S19).

We consider an epidemic to be contained if there are fewer than 500 cases in the 500,000-person community (≤1 per 1000). The containment proportion is the proportion of simulations in which the attack rate is ≤1 per 1000. Another measure of how well we have contained the epidemic is the number of infected people who travel out of the 500,000-person community over the course of the epidemic. If this number is very low or even zero, we have effectively contained spread at the source. The number of cases per 1000 people in the population is another measure of success of the intervention.

Given an initial person infected with the newly emergent influenza strain, there are three possible outcomes: (i) no further people are infected; (ii) there is a small epidemic, between 1 and 500 total cases (≤1 per 1000); or (iii) there is a large epidemic (>1 case per 1000 people). The relative probabilities of these three outcomes as well as the average size of a large epidemic vary with R_0 (Fig. 1C).

Figure 2A shows a typical realization of a large epidemic due to a single initial infective at $R_0 = 1.4$, with no intervention, as well as the average times for intervention initiation. On average, the first symptomatic case appeared 4 days after the initial infection, with intervention initiation times on average 11, 18, or 25 days after the initial infection. Figure 2B shows a typical realization of an epidemic contained with 90% GTAP. Movies 1 to 3 in the SOM show the geographic spread of the epidemic with and without intervention.

Figure 3 gives bar plots of the results for the different intervention strategies and values of R_0. Table 1 gives numbers for the results for R_0 values of 1.4 and 1.7. The measures of containment did not vary much if intervention was initiated 7, 14, or 21 days after the first case, so we give results just for the 14-day delay, followed by a sensitivity analysis of the effect of further delay (fig. S14). When $R_0 = 1.1$, just above threshold, all of the interventions work well. Both 80% TAP and 90% GTAP would be effective in containing pandemic influenza at the source if $R_0 \leq 1.4$. If $R_0 \geq 1.7$, then neither 80% TAP nor 90% GTAP is consistently effective in containing the epidemic, and 300,000 to 350,000 courses of oseltamivir would be needed. Thus, for these interventions singly, a containment threshold exists somewhere between $R_0 = 1.4$ and 1.7. Further sensitivity analysis shows that the containment threshold is roughly at $R_0 = 1.6$.

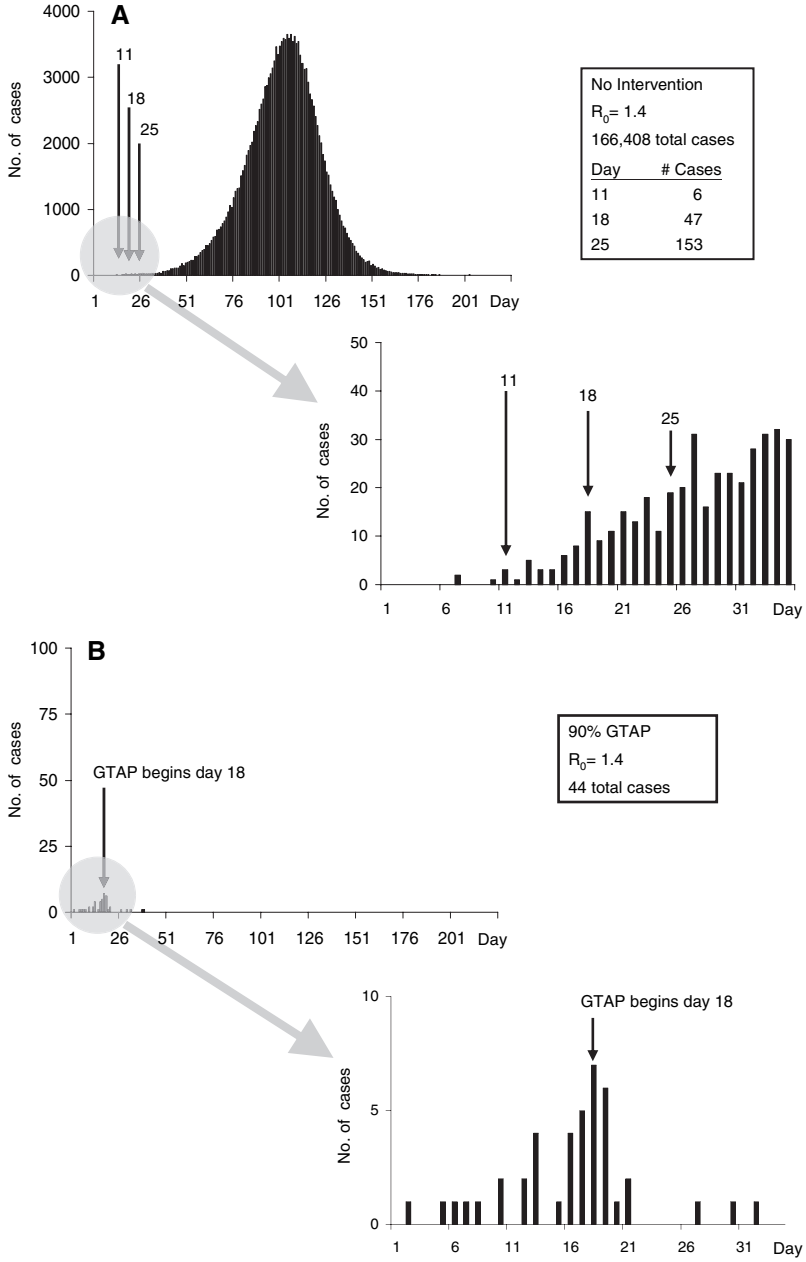


Fig. 2. Model stochastic realizations. (**A**) A typical stochastically simulated large influenza epidemic with no intervention and $R_0 = 1.4$. Also shown are the main intervention initiation times considered and the number of cases at those intervention times. (**B**) A typical stochastically simulated influenza epidemic that is contained using 90% GTAP initiated 14 days after the first case, when $R_0 = 1.4$.

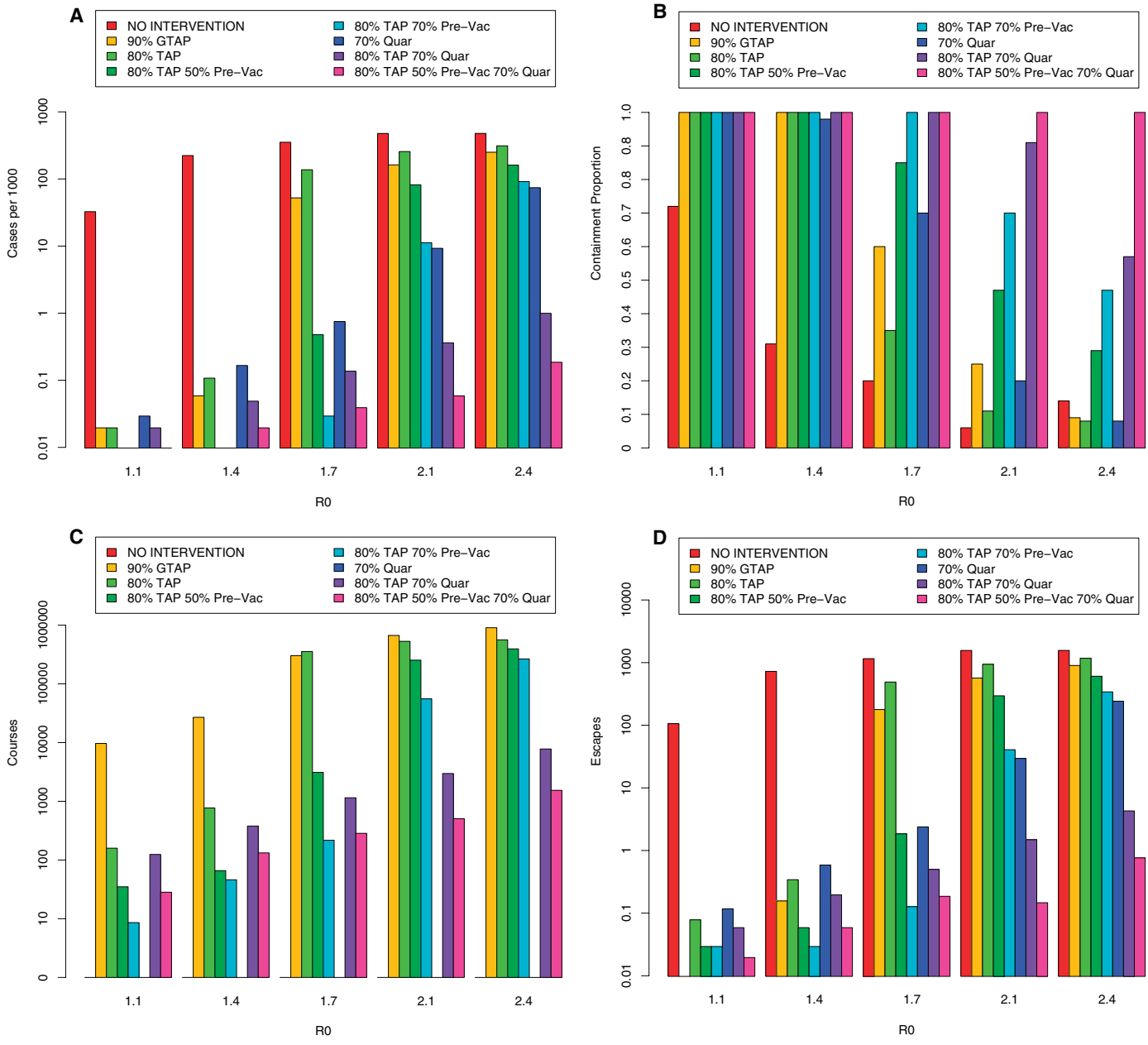


Fig. 3. The effectiveness of the different interventions as compared to no intervention started 14 days after the first case at different values of R_0. The interventions considered are 90% GTAP, 80% TAP, 80% TAP plus 50% pre-vaccination of the population (80% TAP 50% Pre-Vac), 80% TAP plus 70% pre-vaccination of the population (80% TAP 70% Pre-Vac), 70% household and household cluster quarantine (70% Quar), 80% TAP with 70% household and household cluster quarantine (80% TAP 70% Quar), and 80% TAP plus 50% pre-vaccination of the population with 70% household and household cluster quarantine (80% TAP 50% Pre-Vac 70% Quar). (**A**) Average number of cases per 1000 people with no intervention and with different interventions. (**B**) Average containment proportion defined as the proportion of epidemics with one or more secondary cases that had 500 or fewer cases in the population of 500,000 (or $\leq$1 case per 1000). The "no intervention" entry gives the proportion of these epidemics that had 500 or fewer cases with no intervention in the population of 500,000 (or $\leq$1 case per 1000). (**C**) Average number of infected people leaving the 500,000-person population. Each day, the number of infected people who have not withdrawn to the home or are quarantined is multiplied by 10^{-3}, the probability that a person will travel outside of the 500,000-person area. (**D**) Average number of courses of oseltamivir used for the intervention.

Pre-vaccination of the population with a low-efficacy vaccine greatly enhances the effectiveness of TAP and GTAP, even with just 50% coverage. With pre-vaccination, both 80% TAP and 90% GTAP are effective at containing the epidemic when $R_0 = 1.7$, but not at higher levels of R_0. Pre-vaccination essentially lowers the reproductive number (*31*). Local household quarantine is effective at containing the epidemic if $R_0 \leq 2.1$ but is not as effective at $R_0 = 2.4$. However, a combination of 80% TAP plus quarantine is effective at an R_0 as high as 2.4, and adding pre-vaccination makes TAP plus quarantine even more effective.

We conducted a number of sensitivity analyses of the effectiveness of the different interventions at different levels of implementation and delay (figs. S11 to S18). We found that at $R_0 \leq 1.4$, either TAP or GTAP alone is effective only at the 70% level or higher. At higher values of R_0, household quarantine must reach the 70% level to be effective. In terms of timing of the intervention (fig. S14), 90% GTAP and 80% TAP become less effective when the intervention starts 28 days or more after the detection of the first symptomatic case, when there is an average of 85 cases already. Quarantine at the 70% level becomes less effective 42 days (average, 313 cases)

Table 1. Simulated mean cases, escapes, courses, and containment proportion for various interventions and no intervention in a typical rural population of 500,000 people in SE Asia.

Intervention	Cases per 1000		Escapes		Courses		Containment proportion	
	$R_0 = 1.4$	$R_0 = 1.7$	$R_0 = 1.4$	$R_0 = 1.7$	$R_0 = 1.4$	$R_0 = 1.7$	$R_0 = 1.4$	$R_0 = 1.7$
No intervention	211	384	686	1254	–	–	–	–
80% TAP	0.13	149	0.43	525	1,042	381,273	0.98	0.33
90% GTAP	0.28	54	1	187	54,834	325,431	0.95	0.59
80% TAP + 50% pre-vaccination	0.02	0.16	0.06	0.67	87	1,338	1.00	0.98
80% TAP + 70% pre-vaccination	0.01	0.04	0.08	0.12	67	269	1.00	1.00
70% quarantine	0.17	1	0.72	3	-	-	0.98	0.57
80% TAP + 70% quarantine	0.06	0.14	0.18	0.36	484	1,349	1.00	1.00
80% TAP + 70% quarantine + 50% pre-vaccination	0.02	0.03	0.06	0.17	91	275	1.00	1.00

after detection of the first case, even with the addition of TAP. All other interventions in combination with pre-vaccination would be fairly effective even 56 days (average, 894 cases) after the detection of the first case.

It may not be practical to get antiviral agents to exposed people within 1 day of the index case developing symptoms. We carried out a sensitivity analysis for delays in initiation of TAP in the close contact mixing groups ranging from 2 to 5 days after detection of an index case, with 80% TAP. With a delay of up to 2 days, substantial reduction in the number of cases is still achieved, but with delays of 3 to 5 days, there is less benefit (fig. S15). Sensitivity analysis on antiviral efficacy (figs. S16 to S18) shows that the effectiveness of TAP and GTAP is moderately sensitive to variation in AVE_S but not as much to variation in antiviral efficacy in preventing symptomatic disease if infected, AVE_D. Both AVE_S and AVE_I need to be 0.5 or higher for either TAP or GTAP to be effective. Sensitivity analysis on VE_I shows that the effectiveness of 80% TAP with 70% pre-vaccination is sensitive to variation in VE_I (fig. S19). However, even at a level of VE_I as low as 0.1 (fig. S19), the epidemic is still well contained.

We have shown that the targeted use of antiviral agents, if implemented within 21 days of the first case and if $R_0 \leq 1.4$, would have a high probability of success for containing an emergent influenza strain at the source in a rural SE Asian population. Such interventions would be effective for R_0 values as high as 1.7 in the presence of pre-vaccination with a low-efficacy vaccine. For higher values of R_0, localized household quarantine would have to be implemented, possibly in combination with targeted antiviral prophylaxis to contain the pandemic at the source. Although the R_0 of a future newly emergent influenza strain is unknown, previous estimates are 1.89 from the first epidemic of pandemic A (H3N2) in Hong Kong (*19*) and 2 to 3 for 1918 pandemic A (H1N1) in the United States (*32*). However, a newly emergent influenza strain may not yet be well adapted to humans and could have an $R_0 < 2$, and possibly just above 1. As the virus adapts to human-to-human transmission, there would probably be an incremental increase in R_0 with each transmission event (*33*). This makes early intervention especially important.

Based on the results here, the current World Health Organization stockpile of 120,000 treatment courses could possibly be sufficient to contain a pandemic if the stockpile were deployed at the source of the emerging strain within 2 to 3 weeks of detection. Given that early containment at the original source may fail or the emergent strain may appear simultaneously in several locations, up to 1 million courses could be needed to deal with the multiple outbreak foci. In addition, pre-vaccination of populations at risk for a newly emergent influenza strain would be prudent, even if the vaccine provided only moderate protection. Although the effectiveness of most interventions was fairly invariant to the timing of intervention initiation up to 21 days after the first case, delay much beyond that could allow the pandemic to spread unless pre-vaccination takes place.

These results are probabilistic and demonstrate considerable variability in the potential size of the epidemic in the absence of and in response to intervention (*34*). Public health officials need to keep this probabilistic characteristic of success in mind when planning and evaluating their response. We have developed a flexible mathematical model that can help determine the best intervention strategies for containing pandemic influenza at the source. Should a newly emergent influenza strain appear, the model could be quickly calibrated to data and intervention options at the source of the epidemic. Data should be provided from the field to estimate the value of R_0; the serial interval between cases; the distributions of the latent, incubation, and infectious periods; pathogenicity; case fatality ratios; and secondary spread within important mixing groups.

References and Notes

1. R. J. Webby, R. G. Webster, *Science* **302**, 1519 (2003).
2. M. Enserink, *Science* **306**, 2016 (2004).
3. K. Stohr, *N. Engl. J. Med.* **352**, 405 (2005).
4. A. S. Monto, *N. Engl. J. Med.* **352**, 323 (2005).
5. N. M. Ferguson, C. Fraser, C. A. Donnelly, A. C. Ghani, R. M. Anderson, *Science* **304**, 968 (2004).
6. K. Ungchusak *et al.*, *N. Engl. J. Med.* **352**, 333 (2005).
7. K. Stohr, M. Esveld, *Science* **306**, 2195 (2004).
8. F. G. Hayden, *Philos. Trans. R. Soc. London Ser. B* **356**, 1877 (2001).
9. R. G. Webster, *Science* **293**, 1773 (2001).
10. G. Laver, E. Garman, *Science* **293**, 1776 (2001).
11. A. S. Monto *et al.*, *JAMA* **282**, 31 (1999).
12. I. M. Longini, M. E. Halloran, A. Nizam, Y. Yang, *Am. J. Epidemiol.* **159**, 623 (2004).
13. National Statistical Office, *Population and Housing Census 2000* (available at www.nso.go.th, accessed 19 November 2004).
14. K. Faust, B. Entwisle, R. R. Rindfuss, S. J. Walsh, Y. Sawangdee, *Soc. Networks* **21**, 311 (1999).
15. P. Guest, A. Chamraathrithirong, K. Arhavanitkul, N. Piriyathamwong, *Asian Pacific Migrat. J.* **3**, 531 (1994).
16. A. Chamraathrithirong *et al.*, *National Migration Survey of Thailand* (Institute for Population and Social Research, Mahidol Univeristy, Bangkok, Thailand, 1995).
17. L. R. Elveback *et al.*, *Am. J. Epidemiol.* **103**, 152 (1976).
18. E. D. Kilbourne, *The Influenza Viruses and Influenza* (Academic Press, New York, 1975).
19. L. A. Rvachev, I. M. Longini, *Math. Biosci.* **75**, 3 (1985).
20. S. Cauchemez, F. Carrat, C. Viboud, A. J. Valleron, P. Y. Boëlle, *Stat. Med.* **23**, 3469 (2004).
21. W. S. Jordan, *Am. Rev. Res. Dis.* **83**, 29 (1961).
22. I. M. Longini, E. Ackerman, L. R. Elveback, *Math. Biosci.* **38**, 141 (1978).
23. L. E. Davis, G. C. Caldwell, R. E. Lynch, R. E. Bailey, *Am. J. Epidemiol.* **92**, 240 (1970).
24. R. G. Sharrar, *Bull. World Health Org.* **41**, 361 (1969).
25. K. A. Fitzner, S. M. McGhee, A. J. Hedley, K. F. Shortridge, *Hong Kong Med. J.* **5**, 87 (1999).
26. F. G. Hayden, F. Y. Aoki, in *Antimicrobial Therapy and Vaccines*, V. Yu, T. Meigan, S. Barriere, Eds. (Williams and Wilkins, Baltimore, MD, 1999), pp. 1344–1365.
27. F. G. Hayden *et al.*, *N. Engl. J. Med.* **343**, 1282 (2000).
28. R. Welliver *et al.*, *JAMA* **285**, 748 (2001).
29. Y. Yang, I. M. Longini, M. E. Halloran, *Design and Evaluation of Prophylactic Interventions Using Infectious Disease Incidence Data from Close Contact Groups* (Technical Report 04-09, Department of Biostatistics, Emory University, Atlanta, GA, 22 July 2004) (available at www.sph.emory.edu/bios/tech/).
30. M. E. Halloran, I. M. Longini, C. J. Struchiner, *Epidemiol. Rev. Vaccines* **21**, 73 (1999).
31. A. N. Hill, I. M. Longini, *Math. Biosci.* **181**, 85 (2003).
32. C. E. Mills, J. M. Robins, M. Lipsitch, *Nature* **432**, 904 (2004).
33. R. Antia, R. R. Regoes, J. C. Koella, C. T. Bergstrom, *Nature* **426**, 658 (2003).
34. M. E. Halloran, I. M. Longini, D. M. Cowart, A. Nizam, *Vaccine* **20**, 3254 (2002).

35. This work was supported by National Institute of General Medical Sciences MIDAS grant U01-GM070749 and National Institute of Allergy and Infectious Diseases grant R01-AI32042. The authors thank F. Hayden, K. Stöhr, P. Glezen, A. Monto, S. Dowell, and B. Schwartz for their helpful comments.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1115717/DC1
SOM Text
Figs. S1 to S19
Tables S1 to S7
References
Movies S1 to S3

6 June 2005; accepted 14 July 2005
Published online 3 August 2005;
10.1126/science.1115717
Include this information when citing this paper.

In Situ Stable Isotope Probing of Methanogenic Archaea in the Rice Rhizosphere

Yahai Lu[1,2] and Ralf Conrad[2]*

Microorganisms living in anoxic rice soils contribute 10 to 25% of global methane emissions. The most important carbon source for CH_4 production is plant-derived carbon that enters soil as root exudates and debris. Pulse labeling of rice plants with $^{13}CO_2$ resulted in incorporation of ^{13}C into the ribosomal RNA of Rice Cluster I *Archaea* in the soil, indicating that this archaeal group plays a key role in CH_4 production from plant-derived carbon. This group of microorganisms has not yet been isolated but appears to be of global environmental importance.

The translocation of plant photosynthates below ground and their subsequent decomposition by rhizospheric microorganisms are key to the terrestrial ecosystem carbon budget (*1*). It has been shown that 30 to 60% of net photosynthesized carbon is allocated to roots, and as much as 40 to 90% of this fraction enters soil in the forms of root exudates, sloughed-off cells, and decaying roots (*2*). In wetland soils and rice paddies, the below-ground carbon flow provides a major carbon source for methane production (*3*, *4*). However, little is known about the microbiota that are involved in the rhizosphere carbon cycle, and the soil microbiota is considered a "black box" in studies of soil carbon dynamics (*5*).

A combination of stable isotope probing (SIP) and biomarker-based fingerprinting can be a powerful approach to investigate microbial species and in situ activity (*6*, *7*). Biomarkers that have been used with SIP include nucleic acids (*7*–*11*) and phospholipid fatty acids (*6*, *12*–*14*). Nucleic acids are particularly useful because they provide specific taxonomic information (*15*). Here we report on the application of RNA SIP in an intact rice-soil system to identify the active methane producers in the rhizosphere.

Rice fields represent one of the most important individual sources of the greenhouse gas CH_4 (*16*, *17*). Below-ground carbon flow is a key determinant in seasonal patterns of CH_4 emissions from rice fields (*3*). It has previously been shown that pulse labeling of rice plants with $^{14}CO_2$ or $^{13}CO_2$ results in the production and emission of labeled CH_4 from soil (*3*). Furthermore, excised roots can support immediate production of CH_4 if incubated anaerobically (*18*), indicating an active methanogenic microbial community on rice roots and in the rhizosphere (*11*, *19*). These methanogens are differentially activated under various incubation conditions (*11*); however, it is not known how individual phyla contribute to CH_4 production in an intact soil-plant system.

We established duplicate microcosms for ^{13}C labeling experiments (*20*). The microcosms were divided into two chambers to isolate the aerial parts of the plants from the soil and roots. The rice plants grown in each microcosm were repeatedly (49 times over 7 days) pulse-labeled in daylight with $^{13}CO_2$ (99% enrichment) (fig. S1A). The added CO_2 was taken up by the leaves within 1 hour, with the concentration of CO_2 decreasing from 3000 to 100 parts per million by volume during this time. Rice plants were only briefly exposed to higher than ambient CO_2 concentrations, because chronic exposure might eventually have stimulated root exudation and CH_4 production (*21*). CH_4 and CO_2 in soil pore water and CH_4 in the top chamber incorporated ^{13}C rapidly after the assimilation of $^{13}CO_2$ by the plants, with the ^{13}C atom% of CH_4 and CO_2 increasing linearly during the 7-day labeling period (fig. S1B). The value of ^{13}C atom% of CH_4 in the top chamber was less than that of the pore water, because some of the CH_4 produced was presumably oxidized in the rhizosphere (*22*). The rapid ^{13}C labeling of CH_4 in the soil pore water and the top chamber verified that methanogenesis in the rice rhizosphere was highly active and tightly coupled to plant photosynthesis. The ^{13}C atom% of CH_4 in the soil pore water indicated that at least 15% of total CH_4 produced in the rhizosphere originated from plant-derived carbon over the 7-day labeling period (fig. S1B). Because SIP might have resulted in minimal labeling of the rhizosphere microbial community and been barely detectable above the background of natural ^{13}C abundance (*23*, *24*), we established parallel microcosms without ^{13}C labeling as controls (*20*). Nucleic acids from both labeled and nonlabeled microcosms were subjected to the same density-gradient fractionation procedures (*20*). The RNA isolated from labeled rhizospheric soil exhibited, on average, greater buoyant density (BD) (1.796 g ml^{-1}) than that from the nonlabeled rhizospheric soil (BD, 1.788 g ml^{-1}) (fig. S2A). The archaeal templates, as determined by real time polymerase chain reaction using the archaeal-domain specific primers, showed a similar trend (fig. S2B).

Archaeal terminal restriction fragment length polymorphism (T-RFLP) fingerprints (*20*) from RNA taken from different density-gradient fractions after pulse labeling showed a strong shift with increasing BD (Fig. 1A). Specifically, a 394–base pair (bp) fragment markedly increased in frequency in fractions with BD $\geq$ 1.808 g ml^{-1} (Fig. 1, A and B). By contrast, the archaeal fingerprints of non–pulse-labeled soil remained unchanged over all fractions of the density gradient (Fig. 1C). The fingerprints from the "light" RNA fractions of the pulse-labeled soil and those from nonlabeled control soil were almost identical, indicating that the exposure of plants to CO_2 fluctuation caused no bias. We replicated the SIP experiments three times with virtually the same results. Hence, within the archaeal community, the populations with the characteristic 394-bp T-RF had taken up the ^{13}C derived from the pulse labeling and incorporated it into their RNA. The most plausible explanation for this observation is that these archaea were the populations most actively assimilating the carbon provided by the pulse labeling of the rice plants.

To phylogenetically characterize the archaea present in the rhizosphere, we constructed two 16S RNA clone libraries, LrhA (65 clones) and HrhA (62 clones), representing the archaeal light (BD, 1.793 g ml^{-1}) and "heavy" (BD, 1.816 g ml^{-1}) RNA, respectively (*20*). The clone libraries showed that the archaeal populations in the rice rhizosphere consisted of seven dominant lineages (fig. S3), confirming earlier studies on rice roots and rice field soil (*11*, *19*, *25*). *Methanosarcina*-like sequences were predominant, accounting for 55% and 45% of total sequences analyzed for LrhA and HrhA, respectively (Fig. 2). The most obvious difference between the two clone libraries was a much higher abundance of Rice Cluster I (RC-I) in HrhA (26%) than in LrhA (8%). The analysis of the sequence data allowed the assignment of the 394-bp fragment to RC-I and a 187-bp fragment to *Methanosarcinaceae* (fig. S3). Thus, the observation that the 394-bp T-RF increased in the heavy RNA fractions (Fig. 1A) was corroborated with the higher abundance of the RC-I sequences in the HrhA clone library (Fig. 2).

These data demonstrate that, of the archaea detectable in the rice rhizosphere, the RC-I methanogens became specifically labeled. RC-I methanogens have so far not yet been isolated into pure culture; however, an enrich-

[1]College of Resources and Environmental Sciences, China Agricultural University, Beijing 100094, China. [2]Max-Planck-Institute for Terrestrial Microbiology, Karl-von-Frisch-Straße, 35043 Marburg, Germany.

*To whom correspondence should be addressed. E-mail: conrad@staff.uni-marburg.de

ment culture exists (*25*, *26*) that grows using H_2/CO_2 as energy and carbon sources. Hence, it appears that in rice microcosms, photosynthates were excreted from the rice roots and degraded by the anaerobic microbial community. The resulting H_2 allowed the RC-I methanogens to assimilate ^{13}C into ribosomal RNA. The ^{13}C could originate from $^{13}CO_2$ generated by root and microbial heterotrophic respiration or other ^{13}C-phothosynthates that could be assimilated by RC-I methanogens. Our results are consistent with earlier studies on excised rice roots showing that, in this habitat, CH_4 is mainly produced by RC-I methanogens from H_2/CO_2 (*11*, *27*, *28*). Considering the fact that rice fields are an important source for the greenhouse gas methane (*16*, *17*) and that plant-photosynthesized carbon constitutes a major source for CH_4 production in rice soil (*3*, *29*), our study reveals the central importance of the RC-I methanogens in global methane production.

DNA and RNA SIP has frequently been used to link the active microorganisms with the biogeochemical processes operating in various environmental systems (*7*–*11*). The usual disadvantage of the SIP technique is the necessity of adding a ^{13}C-labeled substrate into the environmental sample, which may cause bias by physical disturbance and a change of in situ substrate concentrations. In our experiments, we avoided bias by application of the ^{13}C substrate through pulse labeling of the plant phyllosphere. As rhizosphere carbon flow plays a key role in the interactions between above-ground and below-ground biota (*30*), we believe that RNA SIP has the potential for wide application in rhizosphere microbial ecology, in particular for tracking carbon flow through different compartments of the rhizosphere microbiota and for assessing the effect of environmental conditions on this flow.

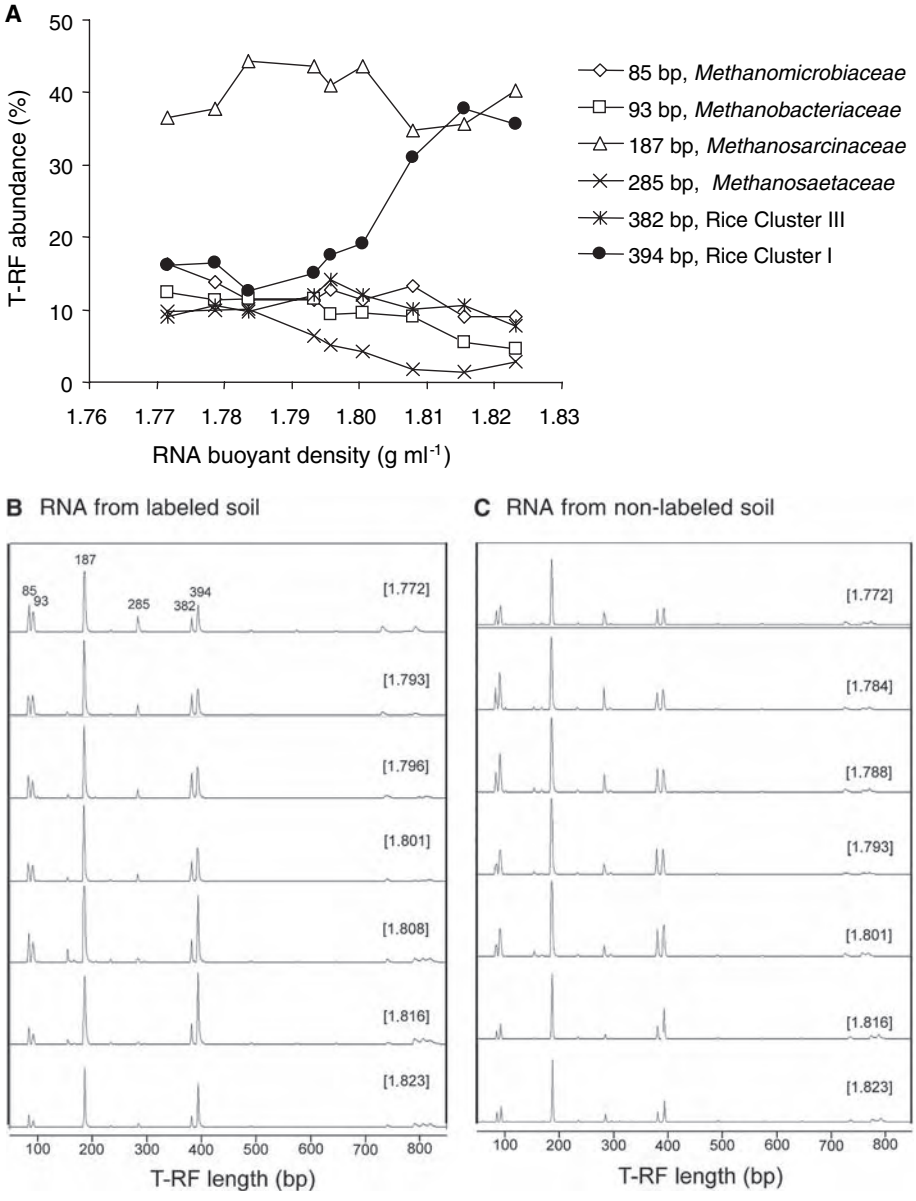


Fig. 1. T-RFLP analysis of archaeal 16S ribosomal RNA from rice rhizospheric soil. (**A**) Relative abundance of different T-RFs across the BD of the isolated ^{13}C-labeled rhizospheric RNA. (**B** and **C**) T-RFLP fingerprints of RNA at different BDs (given in square brackets) from (B) ^{13}C-labeled and (C) nonlabeled rhizospheric soil.

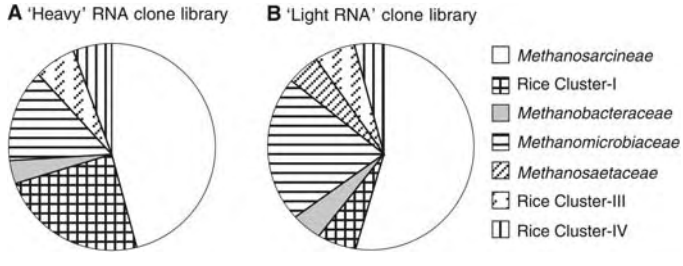


Fig. 2. Phylogenetic affiliation and relative abundance of archaeal 16S ribosomal RNA clones in rhizospheric soil. Relative abundance of the various clones in the (**A**) heavy and (**B**) light RNA clone libraries.

References and Notes

1. P. Hogberg *et al.*, *Nature* **411**, 789 (2001).
2. J. M. Lynch, J. M. Wipps, *Plant Soil* **129**, 1 (1990).
3. T. Minoda, M. Kimura, E. Wada, *J. Geophys. Res.* **101**, 21091 (1996).
4. J. Y. King, W. S. Reeburgh, *Soil Biol. Biochem.* **34**, 173 (2002).
5. K. Paustian, E. Levine, W. M. Post, I. M. Ryzhova, *Geoderma* **79**, 227 (1997).
6. H. T. S. Boschker *et al.*, *Nature* **392**, 801 (1998).
7. S. Radajewski, P. Ineson, N. R. Parekh, J. C. Murrell, *Nature* **403**, 646 (2000).
8. C. O. Jeon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13591 (2003).
9. M. Manefield, A. S. Whiteley, R. I. Griffiths, M. J. Bailey, *Appl. Environ. Microbiol.* **68**, 5367 (2002).
10. T. Lueders, B. Wagner, P. Claus, M. W. Friedrich, *Environ. Microbiol.* **6**, 60 (2004).
11. Y. Lu, T. Lueders, M. W. Friedrich, R. Conrad, *Environ. Microbiol.* **7**, 326 (2005).
12. J. L. Butler, M. A. Williams, P. J. Bottomley, D. D. Myrold, *Appl. Environ. Microbiol.* **69**, 6793 (2003).
13. Y. Lu, J. Murase, A. Watanabe, A. Sugimato, M. Kimura, *FEMS Microbiol. Ecol.* **48**, 179 (2004).
14. A. M. Treonis *et al.*, *Soil Biol. Biochem.* **36**, 533 (2004).
15. B. K. Singh, P. Millard, A. S. Whiteley, J. C. Murrell, *Trends Microbiol.* **12**, 386 (2004).
16. J. Lelieveld, P. J. Crutzen, F. J. Dentener, *Tellus* **50B**, 128 (1998).
17. J. S. Wang *et al.*, *Global Biogeochem. Cycles* **18**, 10.1029/2003GB002180 (2004).
18. R. Conrad, M. Klose, *FEMS Microbiol. Ecol.* **30**, 147 (1999).
19. R. Grosskopf, S. Stubner, W. Liesack, *Appl. Environ. Microbiol.* **64**, 4983 (1998).
20. Materials and methods are available as supporting material on *Science* Online.
21. K. Inubushi *et al.*, *Global Change Biol.* **9**, 1458 (2003).
22. P. L. E. Bodelier, P. Roslev, T. Henckel, P. Frenzel, *Nature* **403**, 421 (2000).
23. R. I. Griffiths *et al.*, *J. Microbiol. Methods* **58**, 119 (2004).
24. J. I. Rangel-Castro *et al.*, *Environ. Microbiol.* **7**, 828 (2005).

25. K. J. Chin, T. Lukow, R. Conrad, *Appl. Environ. Microbiol.* **65**, 2341 (1999).
26. C. Erkel *et al.*, *FEMS Microbiol. Ecol.* **53**, 187 (2005).
27. T. Lueders, K. J. Chin, R. Conrad, M. Friedrich, *Environ. Microbiol.* **3**, 194 (2001).
28. S. Lehmann-Richter, R. Grosskopf, W. Liesack, P. Frenzel, R. Conrad, *Environ. Microbiol.* **1**, 159 (1999).
29. H. A. C. Denier van der Gon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12021 (2002).
30. D. A. Wardle *et al.*, *Science* **304**, 1629 (2004).
31. We thank P. Frenzel for help in the preparation of rice microcosms, M. Klose and P. Claus for laboratory technical assistance, T. Lueders and M. Friedrich for introduction of the SIP technique, and S. Kolb for assistance in phylogenetic analysis. Supported by the Deutsche Forschungsgemeinschaft (grant no. SFB 395). All sequences generated from this study have been deposited in the GenBank database under accession nos. AJ878935 to AJ879061.

Supporting Online Material
www.sciencemag.org/cgi/content/full/309/5737/1088/DC1
Materials and Methods
Figs. S1 to S3
References and Notes

11 April 2005; accepted 6 July 2005
10.1126/science.1113435

Complete Genome Sequence and Lytic Phase Transcription Profile of a *Coccolithovirus*

William H. Wilson,[1]* Declan C. Schroeder,[2] Michael J. Allen,[1] Matthew T. G. Holden,[3] Julian Parkhill,[3] Bart G. Barrell,[3] Carol Churcher,[3] Nancy Hamlin,[3] Karen Mungall,[3] Halina Norbertczak,[3] Michael A. Quail,[3] Claire Price,[3] Ester Rabbinowitsch,[3] Danielle Walker,[3] Marie Craigon,[4] Douglas Roy,[4] Peter Ghazal[4]

The genus *Coccolithovirus* is a recently discovered group of viruses that infect the globally important marine calcifying microalga *Emiliania huxleyi*. Among the 472 predicted genes of the 407,339–base pair genome are a variety of unexpected genes, most notably those involved in biosynthesis of ceramide, a sphingolipid known to induce apoptosis. Uniquely for algal viruses, it also contains six RNA polymerase subunits and a novel promoter, suggesting this virus encodes its own transcription machinery. Microarray transcriptomic analysis reveals that 65% of the predicted virus-encoded genes are expressed during lytic infection of *E. huxleyi*.

Large icosahedral viruses from the family *Phycodnaviridae* infect marine or freshwater eukaryotic algae, and all contain dsDNA genomes ranging from 180 to 560 kbase pair (kbp) (*1*, *2*). Phycodnaviruses belong to a group of viruses that replicate, completely or partly, in the cytoplasm of eukaryotic cells and are termed nucleocytoplasmic large DNA viruses (NCLDVs) (*3*). Two other members of the *Phycodnaviridae* have been sequenced before this study: the 335,593–base pair (bp) genome of a virus that infects a marine filamentous brown alga, *Ectocarpus siliculosus* (EsV-1) (*4*), and the 330,744-bp genome of a virus that infects a unicellular chlorella-like green algal symbiont of the freshwater protozoa *Paramecium bursaria* (PBCV-1). These two viruses, which are from the same family, only have 33 genes in common (*1*).

[1]Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, PL1 3DH, UK. [2]Marine Biological Association, Citadel Hill, Plymouth, PL1 2PB, UK. [3]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [4]Scottish Centre for Genomic Technology and Informatics, Chancellor's Building, College of Medicine, University of Edinburgh, 49 Little France Crescent, Edinburgh EH16 4SB, UK.

*To whom correspondence should be addressed. E-mail: whw@pml.ac.uk

EhV-86, a lytic virus about 170 to 175 nm in diameter, was originally isolated by plaque assay from a seawater sample collected from a dying *Emiliania huxleyi* bloom in the English Channel (*5*). Phylogenetic analysis of its DNA polymerase gene places it in a new genus (*Coccolithovirus*) within the family *Phycodnaviridae* (*6*). The EhV-86 host, *Emiliania huxleyi* (Haptophyta), is a unicellular alga known for its elegant calcium carbonate scales, which it produces intracellularly and sequesters over its cell surface (*7*). It is perhaps best known for its immense coastal and open ocean blooms at temperate latitudes and is a key species for current studies on global biogeochemical cycles and climate modelling (*8*–*10*).

Sequence analysis (*11*) of EhV-86 revealed a circular genome with a length of 407,339 bp (Fig. 1) (*12*), making this the largest *Phycodnaviridae* genome sequenced to date. Other larger algal virus genomes are known to exist, such as the virus that infects *Pyramimonas orientalis*, a marine microalga, which has a genome of about 560 kbp (*13*). Recently, Mimivirus, the largest NCLDV [previously isolated from amoeba growing in the water of a cooling tower (*14*)] was sequenced, revealing a 1,181,404-bp genome (*15*). General features of the EhV-86 genome sequence include a nucleotide composition of 40.2% G+C, a total of 472 predicted genes (coding sequences, CDSs) (Fig. 1) with an average gene length of 786 bp, a coding density of 91%, five tRNA genes [encoding amino acids Leu (containing an intron), Ile, Gln, Asn, and Arg] and two further introns (ehv064 and ehv459). In addition, we identified the location and orientation of three distinct families of repeats (designated A, B, and C) throughout the genome (Fig. 1 and fig. S1). Family A repeats are noncoding but are found immediately upstream of 86 predicted CDSs and are possible promoter elements essential for transcription of associated CDSs. Family B repeats are characterized by GC-rich regions found in CDSs, which encode proline-rich domains. Family C repeats are AT-rich, noncoding, and characteristic of virus genome origins of replication (*16*).

Of the 472 CDSs, only 66 (14%) have been annotated with functional product predictions on the basis of sequence similarity or protein domain matches (Table 1). A large proportion of CDSs exhibited no similarity to proteins in the public databases; only 21% of the CDSs contain protein-protein basic local alignment search tool (BLASTP) results that matched with an *E* value lower than 0.01 (for details of sequence similarity and protein domain database matches, see table S2). Future analysis of marginal similarities may increase this proportion further. Most of the genes with functional predictions are at the "ends" of the genome; surprisingly, there are only 8 such genes within 230 CDSs in the 176-kbp central region between 156 kbp and 332 kbp. The role of this central region is unclear.

Microarray analysis was used to obtain a transcriptome profile during lytic infection and to confirm whether putative CDSs were transcriptionally active. The microarray was constructed from 75-mer oligonucleotides corresponding to the sense strand of each putative CDS on the EhV-86 genome. EhV-86 gene expression (RNA transcription) was determined by hybridization to the microarray with fluorescently labeled cDNAs prepared from an infected *E. huxleyi* culture 33 hours postinfection. This time was chosen because the majority of cells in the culture were infected at a wide range of infection stages (fig. S3); the majority of kinetic classes of transcripts should be expressed at this point, and we detected expression of 308 EhV-86 CDSs

(65%) (Fig. 1 and database S5) (*17*). Analysis of the expression data allowed us to confirm annotation of some overlapping CDSs. Analyses of the transcriptome revealed that many of the unknown CDSs were expressed and are therefore likely to be functional. Up to 35% of EhV-86 CDSs were either not expressed or were beyond the limit of detection at 33 hours postinfection. Interestingly, expression was not observed in a putative phosphate permease (ehv117), a protein predicted to be involved in high affinity phosphate transport (*18*, *19*). Because this study was conducted in nutrient excess, it is possible the permease is only expressed during phosphate-limiting conditions.

Analysis of EhV-86 homologs of known genes revealed several that have never been identified in a virus before. There are at least four genes involved in sphingolipid biosynthesis, encoding sterol desaturase (ehv031), serine palmitoyltransferase (ehv050), transmembrane fatty acid elongation protein (ehv077), lipid phosphate phosphatase (ehv079), and a further two genes encoding desaturases (ehv061 and ehv415). Transcriptomic analysis revealed that all these genes except ehv415 are expressed during infection (Fig. 1). Sphingolipids are membrane lipids present in all eukaryotes and some prokaryotes and also play a key role in several processes, particularly signal transduction (*20*). A sphingolipid biosynthesis pathway has not previously been discovered on a virus genome. Sphingolipid biosynthesis leads to the formation of ceramide (*21*), which is known to suppress cell growth and is an intracellular signal for apoptosis (*22*, *23*). Several viruses are known to induce apoptosis (*24*). Uniquely, EhV-86 appears to encode key components of the ceramide pathway in its genome; furthermore, it is known that there is a connection between protease activation and ceramide-induced apoptosis (*25*), and intriguingly EhV-86 contains eight proteases (including five serine proteases) throughout the genome (ehv021, ehv109, ehv133, ehv151, ehv160, ehv349, ehv361, and ehv447). All proteases except ehv361 are expressed during infection (Fig. 1). One theory is that this algal virus encodes a mechanism for inducing apoptosis as a strategy for killing the host cell and disseminating progeny virions during the infection cycle. Apoptosis, or programmed cell death, has been observed previously in marine phytoplankton (*26*), although usually as a response to nutrient or other physiological stressors (*27*, *28*) rather than virus infection.

Further analysis of the genome reveals that EhV-86 contains 25 of the core set of 40 to 50 conserved virus genes for NCLDVs (*3*) that encode some of the principal features of virion structure and genome replication and expression (Table 1 and table S2). Uniquely for the *Phycodnaviridae*, EhV-86 contains six RNA polymerase subunits (ehv064, ehv105, ehv108, ehv167, ehv399, and ehv434), all of which are expressed. The presence of a putative virus RNA polymerase holoenzyme together with the family A repeats, which we suggest are previously unknown promoters (Fig. 1), would indicate that EhV-86 encodes its own transcription machinery. Hence, expression of some EhV-86 transcripts may occur in the cytoplasm rather than the nucleus (*29*). *Phycodnaviridae* lack RNA polymerases; thus, further phylogenetic analysis may point to coccolithoviruses being reclassified into a distinct subfamily.

EhV-86 is the largest algal virus genome sequenced to date, and it expresses a range of virus genes, revealing functions more commonly observed in animal and plant cells. Viral ceramide production raises questions about cell death in phytoplankton and may have implications for new sources of lipids in the marine food chain. Ongoing definition of the EhV-86 genome, coupled with transcriptomic analysis, will reveal the function of expressed genes that have no homologs in current databases. As more giant viruses are discovered and their genomes sequenced, there is sure to be an explosion of exciting gene discoveries and novel functions.

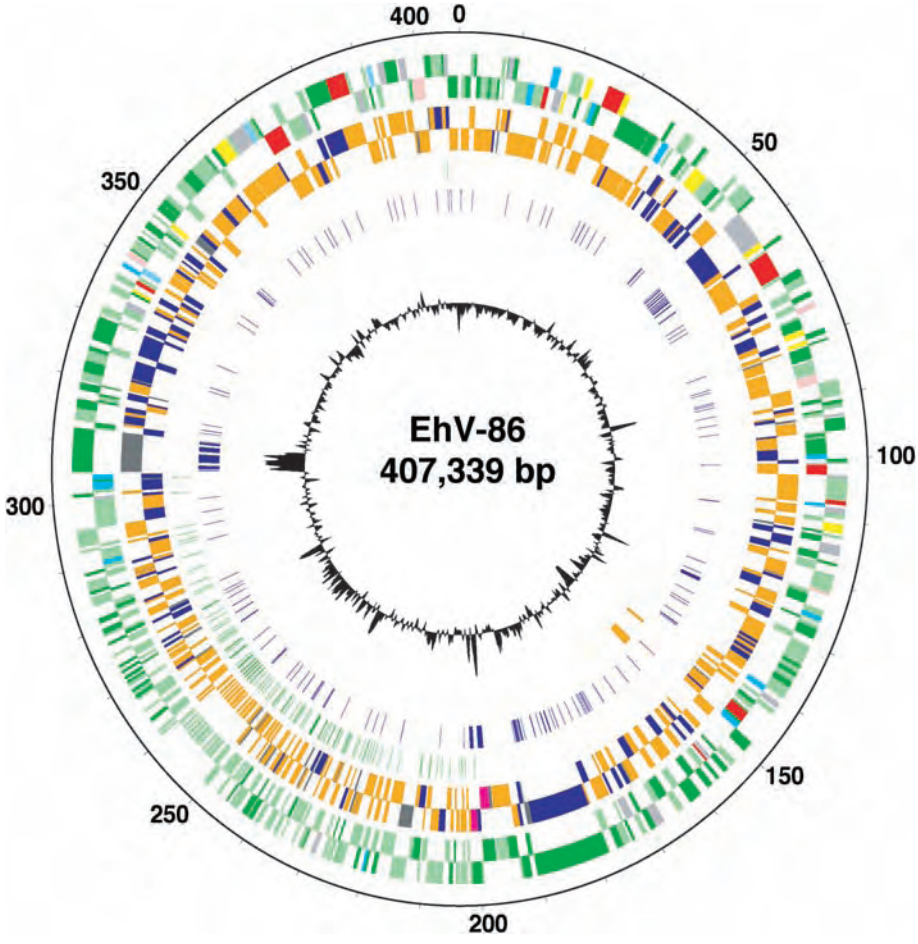


Fig. 1. Circular representation of the 407,339-bp EhV-86 genome. The outside scale is numbered clockwise in kbp. Circles 1 and 2 (from outside in) are CDSs (forward and reverse strands, respectively), starting with CDS ehv001 at position 276 bp (color-coded gray and just below the 0-kb marker on the reverse strand). CDSs are color-coded by putative function: light green, no known function; dark green, no known function but contains transmembrane helices; gray, miscellaneous; sky blue, degradation of large molecules; red, information transfer; yellow, central or intermediary metabolism; pink, virus specific; and light blue, kinases. Circles 3 and 4 are microarray expression results (forward and reverse strands, respectively) color-coded as follows: orange, expression observed; blue, no expression observed; and gray, not determined because oligonucleotides were not designed for that particular CDS. Circles 5 to 7 are the positions of the three families of repeats: green, putative promoter family A; blue, proline-rich family B; and orange, putative origin of replication. Circle 8, G+C content. Initial sequence analysis of EhV-86 suggested a linear genome, because the ends in the sequence alignment always stopped at the same bases. However, polymerase chain reaction amplification across the termini generated a product indicating the genome had a circular stage during virus replication (*30*). Sequence analysis of this product revealed either an A or T (at equal ratio) single base pair overhang at each of the termini, indicating a potential method for circularization of the genome. Consequently, bases T and A were annotated to the "start" of the genome (bases 1 and 2).

Table 1. Genes with functional predictions identified in EhV-86. CDS, systematic gene identifier; CG, CDSs that are known NCLDV core genes (✓) (*3*); Ex, CDSs that are expressed (✓) or not expressed (x) at 33 hours postinfection as determined by microarray transcriptomic analysis.

CDS	Position		Putative function/features	CG	Ex
	Start	End			
ehv014	8433	9299	Longevity-assurance (LAG1) family protein		✓
ehv018	12303	13379	Endonuclease		✓
ehv020	15170	15949	Putative proliferating cell nuclear antigen	✓	✓
ehv021	15998	17212	Serine protease		✓
ehv022	17197	18486	Phosphoglycerate mutase family protein		✓
ehv023	18732	19253	Deoxycytidylate deaminase		✓
ehv026	22027	23004	Ribonucleoside-diphosphate reductase small chain	✓	✓
ehv028	24121	24912	Lipase		✓
ehv030	25675	28713	DNA polymerase delta catalytic subunit	✓	✓
ehv031	28718	29704	Sterol desaturase		✓
ehv041	38172	39305	Endonuclease		✓
ehv050	45226	47838	Serine palmitoyltransferase		✓
ehv060	55942	61926	Lectin protein		x
ehv061	61979	62941	Fatty acid desaturase		✓
ehv064	64018	68501	DNA-dependent RNA polymerase II largest subunit	✓	✓
ehv072	74904	75785	DNA-binding protein	✓	✓
ehv077	78289	79257	Transmembrane fatty acid elongation protein		✓
ehv079	80165	80896	Lipid Phosphate phosphatase		✓
ehv085	86013	87614	Major capsid protein	✓	✓
ehv093	93036	93416	HNH endonuclease family protein		x
ehv101	99368	100174	Hydrolase		x
ehv103	100800	101153	Vesicle-associated membrane protein		✓
ehv104	101158	102741	Putative helicase	✓	✓
ehv105	102837	103331	Transcription factor S-II (TFIIS) family protein	✓	✓
ehv108	106901	107545	DNA-directed RNA polymerase subunit		✓
ehv109	107546	108073	OTU-like cysteine protease		✓
ehv110	108106	108942	RING finger protein		✓
ehv113	110604	112046	Bifunctional dihydrofolate reductase–thymidylate synthase	✓	✓
ehv117	113956	115560	Phosphate permease		x
ehv128	121517	122026	ERV1/ALR family protein	✓	x
ehv133	125567	126283	ATP-dependent protease proteolytic subunit		✓
ehv136	126964	127572	Nucleic acid–binding protein		✓
ehv141	131083	132345	Hypothetical protein	✓	✓
ehv151	139720	140628	Serine protease		✓
ehv152	140747	141610	DNA binding protein		✓
ehv158	145226	147094	DNA ligase	✓	✓
ehv160	147460	148464	Serine protease		✓
ehv166	154991	155719	RING finger protein	✓	x
ehv167	155756	156016	DNA-directed RNA polymerase subunit	✓	✓
ehv179	167551	169176	Major facilitator superfamily protein		✓
ehv184	171775	173091	DNA binding protein		✓
ehv230	219949	220335	Endonuclease		✓
ehv346	228542	229306	Lectin protein		✓
ehv349	289563	290267	Protease		✓
ehv358	300433	300909	Thioredoxin		x
ehv361	302158	303507	Serine protease		x
ehv363	303778	304569	Esterase		x
ehv393	333013	334023	DnaJ domain-containing protein		x
ehv397	335877	336323	Deoxyuridine 5′-triphosphate nucleotidohydrolase	✓	✓
ehv399	337274	337963	DNA-directed RNA polymerase subunit		✓
ehv401	338688	339317	Ribonuclease	✓	✓
ehv402	339307	340335	Protein kinase		x
ehv403	340415	341551	Hypothetical protein	✓	✓
ehv415	348487	349263	Putative fatty acid desaturase		x
ehv428	362962	365202	Ribonucleoside-diphosphate reductase protein	✓	✓
ehv430	365923	368631	Helicase		✓
ehv431	368760	369743	Thymidylate kinase	✓	✓
ehv434	370859	374329	DNA-directed RNA polymerase II subunit	✓	✓
ehv440	378178	379044	Proliferating cell nuclear antigen protein		✓
ehv444	384374	387685	DNA topoisomerase	✓	✓
ehv447	388779	389684	Serine protease		✓
ehv451	391491	392306	Protein kinase	✓	✓
ehv453	393351	394478	MRNA capping enzyme	✓	✓
ehv455	396954	398075	Sialidase		✓
ehv459	398903	401004	Nucleic acid–independent nucleoside triphosphatase	✓	✓
ehv465	404611	405201	Putative thioredoxin protein	✓	x

References and Notes

1. J. L. Van Etten, M. V. Graves, D. G. Muller, W. Boland, N. Delaroque, *Arch. Virol.* **147**, 1479 (2002).
2. W. H. Wilson *et al.*, in *Virus Taxonomy, VIIIth ICTV Report*, C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Dusselberger, L. A. Ball, Eds. (Elsevier/Academic Press, London, 2005), pp. 163–175.
3. L. M. Iyer, L. Aravind, E. V. Koonin, *J. Virol.* **75**, 11720 (2001).
4. N. Delaroque *et al.*, *Virology* **287**, 112 (2001).
5. W. H. Wilson *et al.*, *J. Mar. Biol. Assoc. U.K.* **82**, 369 (2002).
6. D. C. Schroeder, J. Oke, G. Malin, W. H. Wilson, *Arch. Virol.* **147**, 1685 (2002).
7. P. Westbroek, J. R. Young, K. Linschooten, *J. Protozool.* **36**, 368 (1989).
8. R. J. Charlson, J. E. Lovelock, M. O. Andreae, S. G. Warren, *Nature* **326**, 655 (1987).
9. P. Westbroek *et al.*, *Global Planet. Change* **8**, 27 (1993).
10. P. Westbroek *et al.*, in *The Haptophyte Algae*, J. C. Green, B. S. C. Leadbeater, Eds. (Clarendon, Oxford, 1994), vol. 51, pp. 321–334.
11. Materials and methods are available as supporting material on *Science* Online.
12. The complete sequence of EhV-86 is deposited in GenBank/European Molecular Biology Laboratory under accession number AJ890364.
13. R. A. Sandaa, M. Heldal, T. Castberg, R. Thyrhaug, G. Bratbak, *Virology* **290**, 272 (2001).
14. B. La Scola *et al.*, *Science* **299**, 2033 (2003).
15. D. Raoult *et al.*, *Science* **306**, 1344 (2004); published online 14 October 2004 (10.1126/science.1101485).
16. I. Galli, S. M. M. Iguchiariga, H. Ariga, *Nucleic Acids Res.* **20**, 3333 (1992).
17. Microarray data are submitted in the European Bioinformatics Institute ArrayExpress database (www.ebi.ac.uk/arrayexpress) under the accession number e-maxd-2. In addition, this database is also available at the Environmental Genomics Thematic Programme Data Centre (EGTDC) data catalog (http://envgen.nox.ac.uk/) under the accession number egcat:000010.
18. A. Berhe, U. Fristedt, B. L. Persson, *Eur. J. Biochem.* **227**, 566 (1995).
19. A. Berhe, R. Zvyagilskaya, J. O. Lagerstedt, J. R. Pratt, B. L. Persson, *Biochem. Biophys. Res. Commun.* **287**, 837 (2001).
20. A. H. Futerman, Y. A. Hannun, *EMBO Rep.* **5**, 777 (2004).
21. A. H. Merrill Jr., *J. Biol. Chem.* **277**, 25843 (2002).
22. L. M. Obeid, C. M. Linardic, L. A. Karolak, Y. A. Hannun, *Science* **259**, 1769 (1993).
23. Y. A. Hannun, L. M. Obeid, *Trends Biochem. Sci.* **20**, 73 (1995).
24. J. Teodoro, P. Branton, *J. Virol.* **71**, 1739 (1997).
25. S. A. Susin *et al.*, *J. Exp. Med.* **186**, 25 (1997).
26. K. D. Bidle, P. G. Falkowski, *Nat. Rev. Microbiol.* **2**, 643 (2004).
27. C. P. D. Brussaard, A. A. M. Noordeloos, R. Riegman, *J. Phycol.* **33**, 980 (1997).
28. J. A. Berges, P. G. Falkowski, *Limnol. Oceanogr.* **43**, 129 (1998).
29. J. L. Van Etten, *Annu. Rev. Genet.* **37**, 153 (2003).
30. W. H. Wilson *et al.*, unpublished data.

31. This research was supported by a grant in aid from the Natural Environment Research Council (NERC) awarded to the Marine Biological Association (MBA) of the UK and by a grant awarded to W.H.W. from the NERC Environmental Genomics thematic program (ref. NE/A509332/1). We would like to acknowledge support from the EGTDC, Centre for Ecology and Hydrology, Oxford. We thank E. Gudmundsdottir and K. Vierlinger from Scottish Centre for Genomic Technology and Informatics (SCGTI) for help with probe design and initial microarray labeling experiments.

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5737/1090/DC1
Materials and Methods
Figs. S1, S3, S4
Table S2
Database S5

4 April 2005; accepted 5 July 2005
10.1126/science.1113109

Structural Basis for the Activation of Cholera Toxin by Human ARF6-GTP

Claire J. O'Neal,[1]* Michael G. Jobling,[4]* Randall K. Holmes,[4] Wim G. J. Hol[2,3]†

The *Vibrio cholerae* bacterium causes devastating diarrhea when it infects the human intestine. The key event is adenosine diphosphate (ADP)–ribosylation of the human signaling protein $G_{s\alpha}$, catalyzed by the cholera toxin A1 subunit (CTA1). This reaction is allosterically activated by human ADP-ribosylation factors (ARFs), a family of essential and ubiquitous G proteins. Crystal structures of a CTA1:ARF6-GTP (guanosine triphosphate) complex reveal that binding of the human activator elicits dramatic changes in CTA1 loop regions that allow nicotinamide adenine dinucleotide (NAD^+) to bind to the active site. The extensive toxin:ARF-GTP interface surface mimics ARF-GTP recognition of normal cellular protein partners, which suggests that the toxin has evolved to exploit promiscuous binding properties of ARFs.

The *Vibrio cholerae* bacterium causes cholera, a serious diarrheal disease that claims thousands of victims in third-world, war-torn, and disaster-stricken nations each year (*1*). *V. cholerae* secretes its major virulence factor, cholera toxin (CT), when colonizing the mucosa of the human small intestine. CT is composed of a globular A subunit and a pentamer of B subunits (*2*, *3*). Heat-labile enterotoxin (LT) produced by enterotoxigenic *Escherichia coli*, which is responsible for hundreds of thousands of children's deaths from diarrheal diseases annually (*4*), shares more than 80% sequence identity with CT. In both toxins, proteolytic cleavage of the A subunit and reduction of the Cys^{187}-Cys^{199} disulfide bond covalently separate the A1 and A2 domains, formed by residues 1 to 192 and 193 to 240, respectively (*5*). A1 is solely responsible for the toxin's enzymatic activity; however, in vivo delivery of the A1 enzyme to its target requires the nonenzymatic B pentamer and A2 peptide, which testifies to an extensive coevolution of the diarrheal pathogens with their human hosts (*6*, *7*).

A1 is an ADP-ribosyltransferase, catalyzing the covalent transfer of an ADP-ribose moiety from NAD^+ to Arg^{201} of the signaling protein $G_{s\alpha}$. This modification triggers a series of events that culminates in massive efflux of salts and water from the epithelial cells into the intestinal lumen, creating the watery diarrhea characteristic of cholera. ADP-ribosylating toxins (ADPRTs) produced by other bacterial pathogens, as well as human regulatory proteins termed poly(ADP-ribose) polymerases, catalyze similar chemistry with different acceptor residues in different protein targets. Compared with other ADPRTs, A1 by itself has relatively low activity in vitro, but the affinity and enzymatic efficiency of CTA1 toward its substrates is increased by interaction with an ADP-ribosylation factor (ARF) protein from the human host (*8*–*10*). Numerous in vitro studies have demonstrated a tight interaction between A1 and ARFs, and recent work additionally indicated that the A1 fragment of LT interacts with ARF in vivo (*11*), although the essential nature of this interaction is still to be established. ARFs are ~20-kD guanine nucleotide-binding proteins that act as molecular switches essential in membrane trafficking and actin cytoskeleton remodeling in eukaryotic cells (*12*). They have the characteristic G protein fold and bind effector proteins using structural elements, termed switch I, interswitch, and switch II, that differ in conformation between guanosine diphosphate (GDP)–bound and GTP-bound states (*13*–*16*). Only ARF-GTP, not ARF-GDP, can bind human effector proteins and, likewise, only ARF-GTP activates A1 (*8*). Despite extensive study, however, the manner in which ARFs activate A1 has remained a mystery.

To investigate how the toxin recognizes ARF and to understand A1 activation in the presence of ARF, we determined the 1.8 Å crystal structure of a CTA1 variant in complex with human ARF6-GTP. Production of a CTA1:ARF complex was challenging, as CTA1 alone is insoluble and expresses as inclusion bodies (*17*). A plasmid encoding both wild-type CTA1 and ARF6 could not be established because expression of the complex was lethal to cells. However, inactivating CTA1 through mutations of catalytic residues Glu^{110} and Glu^{112} to Asp resulted in a CTA1 variant that could be coexpressed readily with wild-type ARF6 to form a stable, soluble complex that gave well-diffracting crystals (*18*).

The CTA1:ARF6-GTP heterodimer buries an extensive solvent-accessible surface of ~1900 $Å^2$ between the two proteins (Fig. 1). The interface is predominantly hydrophobic, explaining the low solubility of CTA1 in the absence of a binding partner. Polar atoms at the periphery of the interface trap six water molecules, which mediate 8 out of the 11 total hydrogen bonds present in the interface (table S2). ARF6-GTP binds CTA1 using 20 residues, mainly from switches I and II and the interswitch region, that are completely conserved or conservatively substituted among human ARFs (Fig. 1 and fig. S1). Conversely, CTA1 binds ARF6-GTP ~15 Å away from the floor of the active site where many different structural elements meet (Fig. 1). All 23 CTA1 residues that form the ARF-binding surface are completely conserved in LTA1 (*2*), but interface residues in LT-IIa and LT-IIb toxins, which share only ~60% overall sequence identity with A1 but are also activated by ARFs, are not as well conserved (*10*, *19*). Superposing A1 subunits from the CTA1:ARF6-GTP and LT-IIb (*19*) structures shows that, except for N152G, all substituted side chains could maintain similar interactions with partner ARF residues. Together, these results suggest that all A1 orthologs interact with all ARF isoforms in a similar manner.

Our structure shows why ARF6-GDP cannot bind CTA1. The conformations of switch I and switch II in ARF6-GDP would cause steric clashes of the CTA1 α7-α8 loop and α8 helix with switch I, and of CTA1 β4-β5 and β6-β7 loops and the β7 strand with switch II. Additionally, hydrophobic side chains of switch I and interswitch residues Phe^{47}, Val^{49}, Asn^{60}, and Trp^{62}, which contact CTA1, are buried in ARF6-GDP (*20*).

Comparison of the CTA1:ARF6-GTP complex with three structures of ARF-family proteins in complex with human protein effectors, GGA1-GAT:ARF1-GTP (*14*), GRIP domain:ARL1-GTP (*15*), and PDEδ:ARL2-GTP (*16*), reveals three major structural similarities shared among all four complexes.

First, all effectors require ARF or ARF-like protein (ARL) to be GTP-bound. In all cases with known structures (Fig. 2), the vast majority of interactions with the ARF component are made to the switch and interswitch regions, which radically change conformation between GDP- and GTP-bound forms (*12*).

Second, the ARF- or ARL-GTP component maintains a rigid structure independent of binding partner. ARF6-GTP bound to CTA1 is nearly identical to that of isolated ARF6-GTP, with a root mean square deviation (RMSD) of 0.6 Å over all C^α atoms. The structure of ARF6-GTP in complex with CTA1 is similar

Departments of [1]Chemistry and [2]Biochemistry and [3]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA. [4]Department of Microbiology, University of Colorado Health Sciences Center, Aurora, CO 80045, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: wghol@u.washington.edu

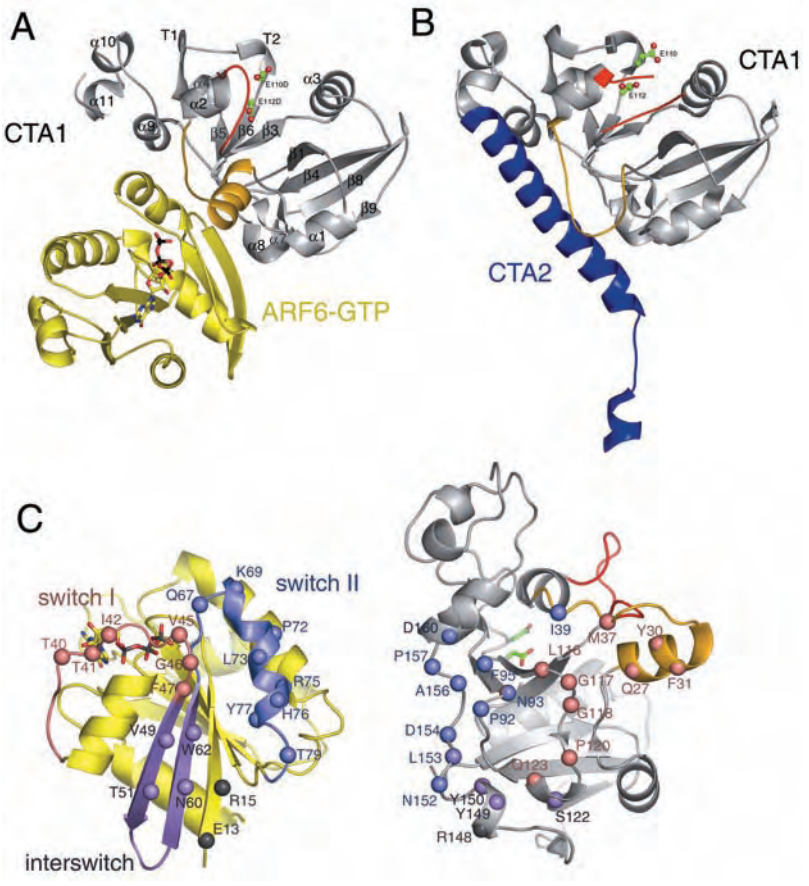


Fig. 1. The complex between bacterial CTA1 and human ARF6-GTP. (**A**) The CTA1:ARF6-GTP heterodimer. Secondary structural elements of CTA1 are labeled according to (*2*). CTA1 is gray, with important loop regions delineated in gold (activation loop) and red (active-site loop). The mutated active-site residues, Asp^{110} and Asp^{112}, are green sticks. ARF6 is yellow, with the bound GTP molecule as sticks. (**B**) CTA1 in complex with CTA2 (blue) from holo-CT. Conformational changes are limited to loop regions, colored as in (A). Active-site loop residue Thr^{50} was disordered in the latent holo-CT structure (*3*); the analogous residue in other inactive holo-CT and LT structures displays increased B factors relative to surrounding residues (*2*, *3*). (**C**) CTA1 and ARF6-GTP interface residues in a "butterfly" representation. Most interactions of CTA1 with ARF6-GTP occur within switch 1 (pink), interswitch (purple), and switch 2 (blue) regions, but hydrogen bonds are also made to N-terminal residues Glu^{13} and Arg^{15} (gray). Each interacting residue is marked with a sphere at the C^{α} position, and CTA1 residues are colored according to the ARF6-GTP region contacted. Abbreviations of amino acid residues as in (*41*).

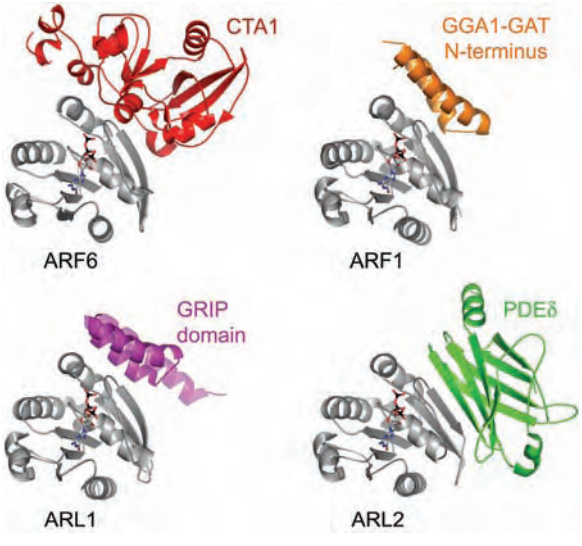


Fig. 2. Comparison of effector binding by ARF- and ARL-GTP. The ARF-family member of each complex is in identical orientation relative to ARF6-GTP from the CTA1:ARF6-GTP structure. Despite greatly varied structures, all effector proteins, including bacterial CTA1, bind a common area on ARF- and ARL-GTP.

to the other effector-bound ARF- and ARLs-GTP, with RMSDs over all C^{α} atoms between 0.6 and 1.1 Å (*14*–*16*).

Third, the interfaces formed between all four effector:ARF-GTP complexes are quite similar. All four effector proteins bind a common ~450 $Å^2$ area on the surface of ARF- or ARL-GTP that is hydrophobic and relatively flat (fig. S2). Each effector:ARF- or ARL-GTP complex structure also recruits three to eight water molecules to the interface periphery to mediate hydrogen bonds between the component proteins (*14*–*16*). Despite the common features, the human complexes vary considerably in buried surface area, from ~1200 to ~1600 $Å^2$, and in shape complementarity, with S_C values ranging from 0.655 to 0.710 (table S3). Interestingly, CTA1 buries more surface area and, with an S_C value of 0.720, has a greater shape complementarity with ARF-GTP than any human effector counterparts.

The similarities between the effector:ARF- or ARL-GTP complexes and CTA1:ARF6-GTP are striking considering that (i) the interaction between human ARF and bacterial CTA1 is an interspecies event and (ii) the different effectors have entirely different structures (Fig. 2). CTA1 has thus evolved to mimic human effector binding to ARF-GTP, which suggests that the multispecific binding strategy of ARF (*21*) apparently leaves it vulnerable to interaction with unintentional partners.

The structure of CTA1 was previously determined in the context of the cholera holotoxin (holo-CT), where it binds CTA2 (*3*, *22*). The structures of CTA2 and ARF6-GTP are completely different (Fig. 1B), yet ~70% of the ARF6-GTP-binding surface of CTA1 also binds CTA2, and conversely, ~50% of the CTA2-binding surface interacts with ARF6-GTP (fig. S3). This is consistent with biochemical results that CTA1 must separate from the holotoxin to interact with ARF-GTP (*23*).

The CTA1 activation loop (residues 25 to 40), earlier suggested to be important in CT and LT activation (*24*), interacts with both partners but in different ways. The loop forms an ordered coil when bound by CTA2 in wild-type holo-CT structures, but residues 25 to 33 rearrange to form an amphipathic helix upon binding ARF6-GTP (Figs. 1 and 3). The helix orients aromatic residues Tyr^{30} and Phe^{31} toward ARF6-GTP switch I, pointing charged side chains Arg^{25}, Asp^{29}, and Arg^{33} to solvent. Mutation of Phe^{31} abolishes interaction between LTA1 and ARF3-GTP (*23*). Interestingly, disorder of activation-loop residues 26 to 36 in a CTA1:Y30S mutant results in a holotoxin variant with intrinsic activity (*3*). Here, activation of CTA1 by ARF6-GTP leads to a well-structured activation loop.

In structures of latent (inactive/uncleaved and unreduced) holo-CT and holo-LT, the active-site loop (residues 47 to 56) occludes

the active site and must move to allow substrate binding (*3*, *24*). When CTA1 is bound to ARF6-GTP, this loop swings out of the active site, with a maximal displacement of 11 Å at Gly^{51} (Figs. 1 and 3). ARF6-GTP binding enables CTA1 active-site loop release through subtle structural shifts near the activation loop (*18*). The open conformation of the active-site loop exposes active-site residues Arg^{7}, Ser^{61}, Glu^{110}, and Glu^{112}, all implicated in substrate binding and catalysis (*25*, *26*). The Glu^{110} and Glu^{112} to Asp mutations in our construct, although inactivating, do not impair the enzyme's ability to bind NAD^+ (*26*). Soaking CTA1:ARF6-GTP crystals with NAD^+ yielded a 2.0 Å crystal structure of the quaternary complex, revealing how this substrate binds to the CTA1 active site (Fig. 4).

Coordination of the ADP moiety of NAD^+ by CTA1 is mostly accomplished by a cluster of arginines (Fig. 4C). The Arg^{7} guanido group coordinates both phosphates, supporting evidence that this residue is essential for NAD^+ binding but is not directly involved in catalysis (*25*). The Arg^{11} guanido group makes van der Waal's interactions with the adenine ribose and π stacks with the adenine base. The Arg^{25} guanido group, pointed away from the active site in the apoprotein complex structure, swings over the bound dinucleotide to hydrogen-bond to the ribose O_A4' and the adenine phosphate and forms van der Waals interactions with the adenine base. Mutations of Arg^{11} or Arg^{25} deleteriously affect toxin activity and toxicity (*27*).

The nicotinamide mononucleotide (NMN) moiety is bound in a compact, energetically unfavorable conformation similar to that seen in other ADPRT-NAD^+ complexes (Fig. 4 and fig. S3) (*28*–*33*). In CTA1, the nicotinamide ring packs between Val^{72} from the α3 helix on one face and backbone atoms from Ser^{61}, Thr^{62}, and Ser^{63} of the conserved "STS (Ser Thr Ser) motif" (*34*) on the other face. Most other ADPRTs stack the ring against a Tyr or Phe residue instead of Val, which may contribute to their greater affinity for NAD^+ relative to CT (*35*). The carboxamide group on the nicotinamide is coordinated in a bidentate fashion by the carbonyl and amide backbone groups of one residue, Ala^{8}, a feature conserved among all ADPRT:NAD^+ complexes.

The nicotinamide ribose ring itself interacts with the mutated residues Asp^{110} and Asp^{112}. Glu^{112} has been implicated by numerous biochemical studies as the catalytic residue absolutely required for activity (*25*, *26*). Here, the Asp^{112} carboxylate is within hydrogen-bonding distance (3.2 Å) of the O_N2' of the nicotinamide ribose ring. The Asp^{110} carboxylate is oriented out into the active site over the ribose ring hydroxyls, within Coulombic-interaction distance of the reaction center. The exposed position of Asp^{110}, together with the essential nature of this residue, suggests that Glu^{110} coordinates and possibly deprotonates the incoming Arg^{201} from $G_{S\alpha}$, improving its nucleophilicity for attack at the C_N1' of NAD^+ (*26*, *36*–*38*).

Certain ADPRTs share two crucial features near the active site despite limited sequence and structural homology: the active-site loop and the ADP-ribosylating turn-turn (ARTT) motif (*39*). In diphtheria toxin and *Pseudomonas aeruginosa* ExoS and ExoT, the active-site loop is required to recognize host-protein substrates and changes conformation to allow NAD^+ access to the active site (*28*, *39*). In CTA1:ARF6-GTP, the open active-site loop creates a new "knob" near the active site from loop residues Thr^{48} to Phe^{52} (Fig. 4B). This

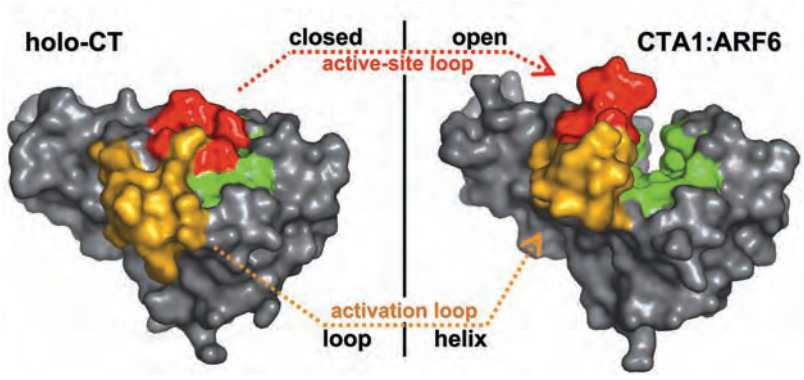


Fig. 3. Changes in loop regions of CTA1 between holo-CT (left) and CTA1:ARF6-GTP (right) lead to opening of the active site (green). Loops are colored as in Fig. 1.

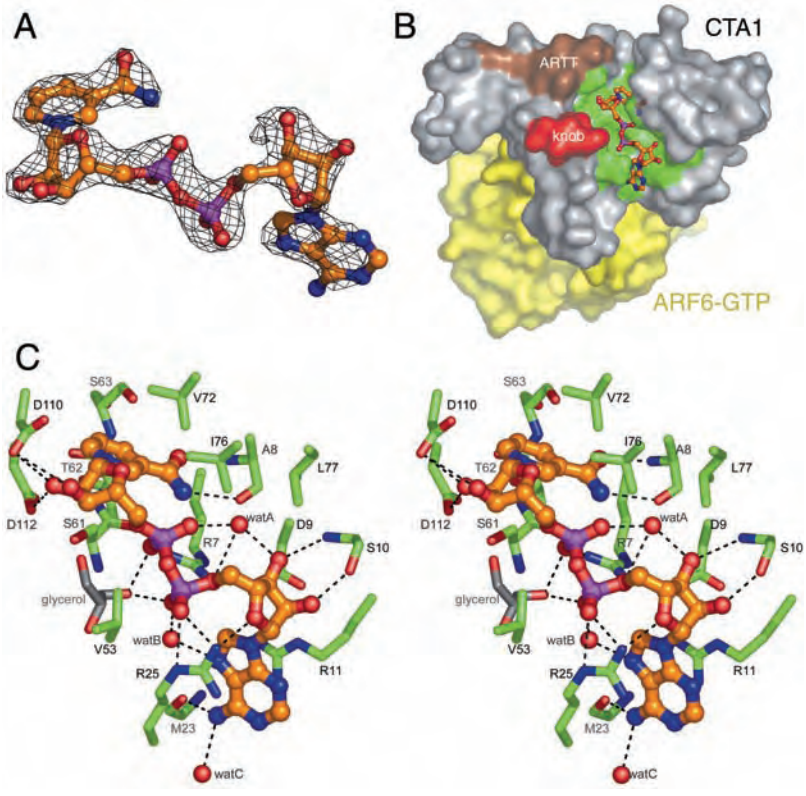


Fig. 4. ARF6-GTP-bound CTA1 in complex with its substrate, NAD^+. (**A**) A σ_A-weighted omit map contoured at 2.5 σ, showing density surrounding the NAD^+ molecule in the CTA1 active site. (**B**) Surface representation of the NAD^+-occupied CTA1 active site (green), viewed from the top. The large active-site cleft is open to solvent from the top and from both ends. ARF6-GTP, in yellow, binds to CTA1 (gray) far from the active site (green). The knob (red), formed by active-site loop residues 48 to 52 when CTA1 is ARF-bound, and the ARTT motif (brown) (*40*) form a surface for potential $G_{S\alpha}$ recruitment. (**C**) Stereoview of contacts between NAD^+ and active site residues (green sticks). Hydrogen bonds are indicated as black dashes. A detailed listing of contacts between NAD^+ and CTA1 atoms is presented in table S4. NAD^+ makes extensive direct interactions with protein atoms but also recruits three waters and a glycerol molecule to mediate hydrogen bonds with the active site. Abbreviations of amino acid residues as in (*41*).

knob forms atop CTA1 near the ARTT motif (residues 104 to 110), which is positionally conserved among ADPRTs and has been shown to play a role in target protein recognition (*40*). Although interaction between the CTA1 active-site loop and $G_{s\alpha}$ has not been studied, these results suggest that the activating human G protein, ARF, alters the bacterial CTA1 structure such that the toxin adopts an enhanced conformation for attacking its human G-protein substrate, $G_{S\alpha}$.

References and Notes

1. D. A. Sack, R. B. Sack, G. B. Nair, A. K. Siddique, *Lancet* **363**, 223 (2004).
2. T. K. Sixma *et al.*, *Nature* **351**, 371 (1991).
3. C. J. O'Neal, E. I. Amaya, M. G. Jobling, R. K. Holmes, W. G. J. Hol, *Biochemistry* **43**, 3772 (2004).
4. R. Widdus, *Bull. World Health Organ.* **79**, 713 (2001).
5. J. J. Mekalanos, R. J. Collier, W. R. Romig, *J. Biol. Chem.* **254**, 5855 (1979).
6. L. De Haan, T. R. Hirst, *Mol. Membr. Biol.* **21**, 77 (2004).
7. J. Hacker, J. B. Kaper, *Annu. Rev. Microbiol.* **54**, 641 (2000).
8. R. A. Kahn, A. G. Gilman, *J. Biol. Chem.* **261**, 7906 (1986).
9. M. Noda, S. C. Tsai, R. Adamik, J. Moss, M. Vaughan, *Biochim. Biophys. Acta* **1034**, 195 (1990).
10. C. M. Lee *et al.*, *J. Clin. Invest.* **87**, 1780 (1991).
11. X. Zhu, R. A. Kahn, *J. Biol. Chem.* **276**, 25014 (2001).
12. S. Pasqualato, L. Renault, J. Cherfils, *EMBO Rep.* **3**, 1035 (2002).
13. S. R. Sprang, *Annu. Rev. Biochem.* **66**, 639 (1997).
14. T. Shiba *et al.*, *Nat. Struct. Biol.* **10**, 386 (2003).
15. B. Panic, O. Perisic, D. B. Veprintsev, R. L. Williams, S. Munro, *Mol. Cell* **12**, 863 (2003).
16. M. Hanzal-Bayer, L. Renault, P. Roversi, A. Wittinghofer, R. C. Hillig, *EMBO J.* **21**, 2095 (2002).
17. L. Agren, M. Norin, N. Lycke, B. Lowenadler, *Protein Eng.* **12**, 173 (1999).
18. Materials and methods are available as supporting material on *Science* Online.
19. F. van den Akker *et al.*, *Structure* **4**, 665 (1996).
20. J. Menetrey, E. Macia, S. Pasqualato, M. Franco, J. Cherfils, *Nat. Struct. Biol.* **7**, 466 (2000).
21. V. Biou, J. Cherfils, *Biochemistry* **43**, 6833 (2004).
22. R. G. Zhang *et al.*, *J. Mol. Biol.* **251**, 563 (1995).
23. X. Zhu *et al.*, *Biochemistry* **40**, 4560 (2001).
24. F. van den Akker *et al.*, *Biochemistry* **34**, 10996 (1995).
25. Y. Lobet, C. W. Cluff, W. Cieplak Jr., *Infect. Immun.* **59**, 2870 (1991).
26. W. Cieplak Jr., D. J. Mead, R. J. Messer, C. C. Grant, *J. Biol. Chem.* **270**, 30545 (1995).
27. M. G. Jobling, R. K. Holmes, *J. Bacteriol.* **183**, 4024 (2001).
28. C. E. Bell, D. Eisenberg, *Biochemistry* **35**, 1137 (1996).
29. H. R. Evans *et al.*, *J. Biol. Chem.* **278**, 45924 (2003).
30. S. Han, J. A. Craig, C. D. Putnam, N. B. Carozzi, J. A. Tainer, *Nat. Struct. Biol.* **6**, 932 (1999).
31. M. Li, F. Dyda, I. Benhar, I. Pastan, D. R. Davies, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6902 (1996).
32. J. Menetrey *et al.*, *J. Biol. Chem.* **277**, 30950 (2002).
33. H. Tsuge *et al.*, *J. Mol. Biol.* **325**, 471 (2003).
34. G. Masignani, M. Pizza, R. Rappuoli, in *Bacterial Protein Toxins*, K. Aktories, I. Just, Eds. (Springer, Berlin, 1999), vol. 145, pp. 21–44.
35. G. C. Zhou *et al.*, *J. Am. Chem. Soc.* **126**, 5690 (2004).
36. C. Wilde, I. Just, K. Aktories, *Biochemistry* **41**, 1539 (2002).
37. M. Nagahama, Y. Sakaguchi, K. Kobayashi, S. Ochi, J. Sakurai, *J. Bacteriol.* **182**, 2096 (2000).
38. S. Han, A. S. Arvai, S. B. Clancy, J. A. Tainer, *J. Mol. Biol.* **305**, 95 (2001).
39. J. Sun, A. W. Maresso, J. J. Kim, J. T. Barbieri, *Nat. Struct. Mol. Biol.* **11**, 868 (2004).
40. S. Han, J. A. Tainer, *Int. J. Med. Microbiol.* **291**, 523 (2002).
41. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

42. We thank N. H. Greene for assistance with crystallization, F. Athappilly and M. A. O'Neal for maintenance of computational facilities, and D. R. Davies, B. E. Krumm, E. Fan, and C. L. M. J. Verlinde for helpful discussions. We are indebted to G. Wisedchaisri and J. Abendroth for assistance with data collection. Funding for this study was provided by NIH grants AI-34501 (W.G.J.H.) and AI-31940 (R.K.H.) and by a Howard Hughes Medical Institute predoctoral fellowship to C.J.O. Coordinates and structure factors for the CTA1:ARF6-GTP, CTA1-NAD^+:ARF6-GTP, and CTA1:ARF6(Q67L)-GTP structures have been deposited in the Protein Data Bank under accession codes 2A5D, 2A5F, and 2A5G, respectively.

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5737/1093/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 to S4
References

11 April 2005; accepted 5 July 2005
10.1126/science.1113398

Pre-Unfolding Resonant Oscillations of Single Green Fluorescent Protein Molecules

Giancarlo Baldini,* Fabio Cannone, Giuseppe Chirico

Fluorescence spectroscopy of a green fluorescent protein mutant at single-molecule resolution has revealed a remarkable oscillatory behavior that can also be driven by applied fields. We show that immediately before unfolding, several periodic oscillations among the chemical substates of the protein chromophore occur. We also show that applied alternating electric or acoustic fields, when tuned to the protein characteristic frequencies, give rise to strong resonance effects.

Proteins exhibit a large number of conformational substates that are associated with a complex energy landscape. At physiological temperature, internal specific motions enable a protein to sample its energy landscape by frequent jumps among the different states. This conformational flexibility appears to be intimately connected to protein reactivity; examples include heme proteins during the kinetics of ligand binding (*1*, *2*), enzyme-substrate dynamics (*3*), and folding of RNA metastable intermediates (*4*). In the past decade, attention has been extended to the internal dynamics of various biomolecules by fluorescence resonance energy transfer of engineered proteins (*5*, *6*) and by atomic force microscopy (*7*). Classical investigations on biosystems have been performed mainly on ensembles that contain extremely large numbers of molecules; consequently, conformational distributions have only been indirectly inferred through models, as experiments have yielded data averaged over all the states simultaneously.

It is thus expected that the nature, the dynamics, and the role of the conformational substates can be successfully unraveled by single-molecule experiments, provided that sufficiently high space and time resolution is available. Improved experimental techniques have recently shown great promise in revealing single-molecule dynamics (*8*–*11*). Fluorescence spectral fluctuations, for example, could be attributed to specific enzyme-substrate dynamics (*4*, *12*). Another study revealed the protein conformational kinetics underlying fluorescence fluctuations in staphylococcal nuclease molecules (*13*). Single-molecule techniques have also been used to investigate protein folding and unfolding (*5*, *6*, *14*). In addition, theoretical predictions of fluorescence fluctuations associated with oscillations among different chemical states of a single molecule have been published (*15*).

Transitions among the conformations of green fluorescent protein (GFP), a bioluminescent protein from the jellyfish *Aequorea victoria*, have been widely investigated both experimentally (*16*) and theoretically (*17*). The intrinsic fluorescence emission of GFP has received considerable attention for its applications to molecular biology and biotechnology (*18*, *19*). The GFP chromophore can adopt four distinct chemical substates: anionic A (deprotonated), neutral N (protonated), intermediate I (chemically similar to A), and zwitterionic Z (nonfluorescent) (*20*–*22*). Several photoconversion (fig. S1) pathways among the A, N, and I substates have been established (*21*, *23*, *24*). Under laser illumination, the GFP fluorescence displays random switching between A and N substates in the microsecond to millisecond range (*24*–*27*).

When dealing with single proteins, the contributions of each conformational substate

Laboratory for Advanced BioSpectroscopy, Physics Department, and Centro Nazionale delle Ricerche (CNR)–Istituto Nazionale per la Fisica della Materia (INFM), University of Milano–Bicocca, Milano I-20126, Italy.

*To whom correspondence should be addressed. E-mail: baldini@mib.infn.it

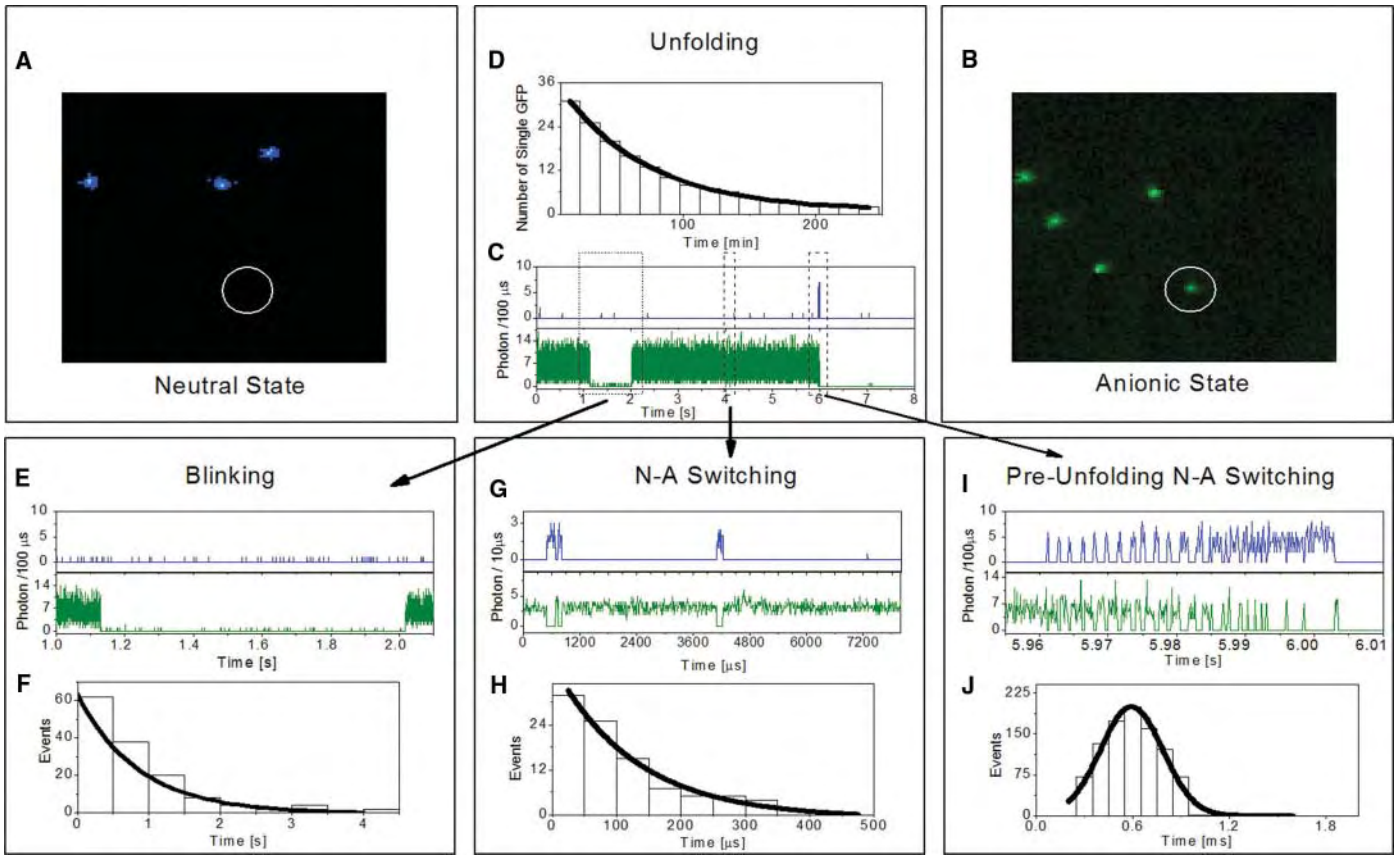


Fig. 1. (**A** and **B**) Two-photon fluorescence images of eight single GFPs in the 8 μm by 8 μm field of view (magnification, ×100; residence time per pixel, 1 ms). Blue spots are molecules in the neutral state N; green spots are molecules in the anionic state A. (**C**) GFP fluorescence time evolution of the single GFP circled in the images after the addition of 4.45 M GuHCl 1 hour earlier. Data were simultaneously acquired in the neutral (blue) and anionic (green) channels at sampling time = 100 μs. The data for the first 3593 s are not shown here. (**D**) Number of fluorescent GFPs versus time; the solid line is an exponential fit to the data with average unfolding time ≈ 64 min. Three time windows in (C) are expanded in the following panels. (**E**) Blinking in the 1.0- to 2.1-s window [same color code as in (C)]. (**F**) Blinking Δt_{off} distribution of 10 molecules; the solid line is an exponential fit to the data ($\langle \Delta t_{off} \rangle_B$ = 860 ± 80 ms). (**G**) Fluorescence trace in the 4.000- to 4.008-s window [same color code as in (C); sampling time = 10 μs]. (**H**) Anionic fluorescence Δt_{off} distribution of 10 molecules; the solid line is an exponential fit to the data ($\langle \Delta t_{off} \rangle_S$ = 140 ± 20 μs). (**I**) N-A switching observed in the 5.955- to 6.010-s time window (i.e., just before the unfolding event). (**J**) Anionic fluorescence Δt_{off} distribution of 10 molecules; the solid line is a Gaussian fit to the data ($\langle \Delta t_{off} \rangle_O$ = 0.64 ± 0.24 ms standard deviation).

may be difficult to resolve because interconversion kinetics could be much too fast, the excitation probability could be much too low, and/or the signals coming from the various substates could be too close in frequency to allow the spectral discrimination. GFP, however, appears an ideal candidate for investigating the transitions among conformations because it is easy to discern the different fluorescence emissions from its well-characterized states (*23*). In particular, the A state fluoresces in the green and the N state in the blue. Moreover, in the proximity of unfolding, the role of some conformational transitions may be amplified as a result of structure relaxation.

These aspects have led us to study a protein mutant, GFPmut2 (*28*, *29*), as a suitable system for investigating the dynamics of conformation interconversion close to unfolding induced by guanidium hydrochloride (GuHCl) and other denaturants (*30*). To check consistency between single-molecule and ensemble measurements, we compared the unfolding rate and the native-denatured equilibrium in the two conditions. Single GFPmut2 molecules trapped in silica gels, monitored by fluorescence, have given data in excellent agreement with fluorescence and circular dichroism from bulk solutions (*31*, *32*). Encapsulation of proteins in silica gels (*33*, *34*) is a very effective method of enhancing their stability (*35*, *36*) so as to isolate and characterize individual conformations (*37*) and to single out internal protein dynamics (*38*). Water-rich matrices such as the gels used here offer a good simulation of the ambient cell, and (unlike solutions) they allow observation of the molecule for several hours, as required when recording repeated conversions. The long-term stability of the protein has been demonstrated by fluorescence measurements that show the unfolding-refolding process of GFPmut2 to be fully reversible over several repeated cycles. The data have also been reproduced in other matrices and with urea as a denaturant (fig. S6).

In Fig. 1 we report typical two-photon fluorescence images of a gel that hosts isolated GFP molecules (*29*) recorded with the microscope apparatus described in (*30*). Each spot corresponds to a single GFP either in the anionic state A (green spots) or in the neutral state N (blue spots). After each of the several induced unfolding-refolding cycles (fig. S2), the molecules are found in either the A or the N state independent of the state displayed in the previous unfolding. Fluorescence, recorded simultaneously in both channels (Fig. 1), shows signals that can last hours, depending on denaturant concentration (*31*), followed by a sudden signal loss that corresponds to unfolding. The recorded fluorescence kinetics of single GFP proteins indicate the following photodynamic processes: unfolding from either state, A/N → unfolded state (Fig. 1C); blinking, A ↔ Z (Fig. 1E); and switching, A ↔ N (Fig. 1G). The unfolding events occur with an average delay time of 64 ± 5 min after the addition of 4.45 M GuHCl (Fig. 1D), and they are monitored by a sudden loss of fluorescence that can be recovered only by removing the denaturant. Fluorescence fluctuations can be described by on- and off-times

(*26*, *39*). Blinking corresponds to a temporary [$\langle \Delta t_{off} \rangle_B$ = 1.0 ± 0.2 s (Fig. 1F)] (*26*, *29*) and reversible loss of fluorescence from the A channel only. Switching or photoconversion, unlike unfolding, is a random and fast fluctuation between the N and A states with $\langle \Delta t_{off} \rangle_S$ = 130 ± 20 μs (Fig. 1H) (*23*, *27*).

Surprisingly, when the data close to the unfolding event are displayed at higher time resolution, several regular A ↔ N transitions appear (Fig. 1I). Because data are acquired simultaneously in both the A and N channels, and fluorescence is found in either channel but never in both at the same time, it appears that each molecule, before unfolding, oscillates several times between the two states. This pre-unfolding oscillatory regime, displayed by all 300 of the investigated single GFP molecules, has a duration $\Delta\tau_0 \approx 42 \pm 1.9$ ms at 4.45 M GuHCl (fig. S3), after which the fluorescence signal vanishes abruptly. The value of $\Delta\tau_0$ decreases slightly as GuHCl concentration increases ($\Delta\tau_0$ = 46.8 ± 2.4 ms at 0.45 M, $\Delta\tau_0$ = 39 ± 1.6 ms at 6.0 M GuHCl). The features of the pre-unfolding kinetics do not depend on whether the cycle begins in the A state or in the N state.

Closer inspection of the fluorescence signal shows that the switching events in the initial part of the pre-unfolding trace are much more regular than those in the final section (in the first part, the spread of the distribution of the switching period ΔP is less than 10%; fig. S4G). The increased regularity of the switching is supported by the Δt_{off} distribution, which is well described by a Gaussian (Fig. 1J), contrary to the trend of all other photodynamic events (Fig. 1, F and H), which are well represented by exponential distributions. Also, ΔP (*39*) exhibits a Gaussian distribution (Fig. 2), suggesting the appearance of a new periodic process. The transition from random to periodic behavior is preceded by a rapid increasing in the switching frequency, and finally it occurs only in the 20- to 40-ms time span before unfolding (Fig. 2D). A ΔP distribution analysis of the first part of the pre-unfolding fluorescence trace (Fig. 2, A to C) of nearly 300 GFPmut2 molecules yields $\langle \Delta P \rangle$ = 2.25 ms (corresponding to frequency ν_1 = 443 Hz) at low denaturant concentration (0.45 M). At higher GuHCl concentrations, other peaks, $\langle \Delta P \rangle$ = 1.38 ms (corresponding to ν_2 = 755 Hz) and $\langle \Delta P \rangle$ = 1.03 ms (corresponding to ν_3 = 971 Hz), appear as well. The observed ΔP distributions were fitted to three Gaussians that were found to bear similar standard deviations ($\sigma_0 \approx 0.16$ ms) independent of GuHCl concentration.

Each unfolding molecule displays one single oscillation frequency, independent of the laser excitation intensity (table S1), and no memory of the frequency assumed in the previous unfolding seems to remain (fig. S5). We have also found that inhomogeneities of the denaturant concentrations in the trapping matrix do not seem to have a role in the observed behavior. Even when the protein is trapped in other environments [such as polyelectrolyte films (fig. S6), agarose gel, and nanocapsules (*40*)], or when different denaturant agents such as urea are used, the oscillation frequencies do not change; they appear to be a specific property of the GFP. ΔP analysis of the whole fluorescence trace that includes all the oscillations, instead of the early part only (as above), shows that the ΔP distribution flattens out as expected from the irregular ending of the switching (fig. S4).

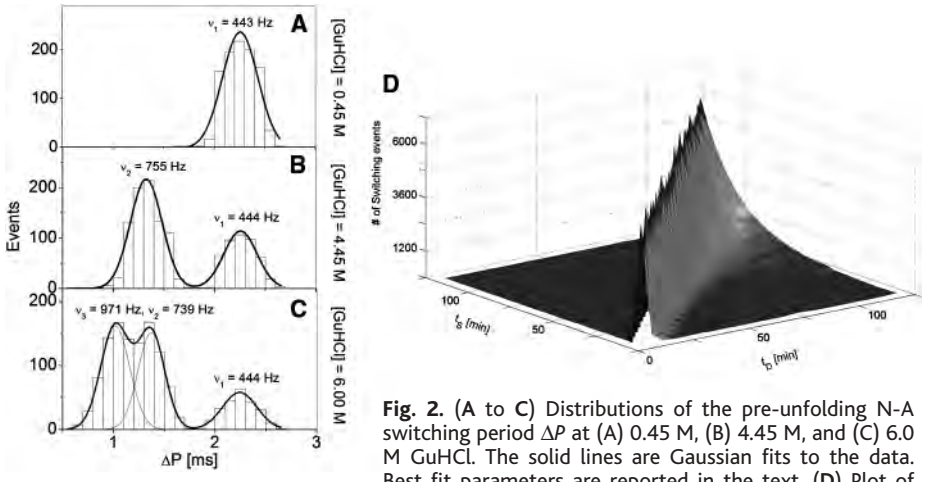


Fig. 2. (**A** to **C**) Distributions of the pre-unfolding N-A switching period ΔP at (A) 0.45 M, (B) 4.45 M, and (C) 6.0 M GuHCl. The solid lines are Gaussian fits to the data. Best fit parameters are reported in the text. (**D**) Plot of number of switching events (computed on intervals of 5 min each) versus switching time t_S (time at which the events are recorded) and unfolding time t_D. Data are from 30 molecules at 0.45 M GuHCl. After t_D the fluorescence vanishes because of molecule unfolding.

We next attempted to drive oscillations with the use of external ac electric and acoustic fields. Electric fields at frequency ν_a, $E(\nu_a)$, applied with platinum electrodes placed in the silica gel containing the molecules, gave marked effects only when the molecular frequencies ν_i fell within the bandwidth $\Delta\nu$ of the field. Modulated acoustic fields, $S(\nu_a)$, coming from earphones placed close to the gel surface, yielded similar results. Neither of the applied fields caused any observable (less than 50 nm) shift of the molecules from their unperturbed location.

Figure 3 shows examples of repeated unfolding-refolding cycles for two GFPs randomly chosen from 102 investigated molecules. The number of observed fluorescence oscillations became substantially larger as the applied frequency ν_a approached one of the characteristic frequencies ν_i (Fig. 3, A, C, D, and F), whereas no such effect resulted if the field frequency was more than one bandwidth, σ_0, away from ν_i (Fig. 3, B and G). Moreover, no driving was observed when the protein oscillated at one particular frequency ν_i and was stimulated at $\nu_j \neq \nu_i$ (Fig. 3H). At resonance under either field (i.e., $\nu_a \approx \nu_i$), the duration of fluorescence oscillation doubled, $\sim 2\Delta\tau_0$ (Fig. 3, A, C, D, and F), and regularity improved considerably. ΔP distribution analysis of the first part of the oscillating fluorescence trace yielded less noisy and sharper ΔP distributions with widths of $\sim 0.5\sigma_0$ (Fig. 4, dashed line) for the GFPs oscillating at resonance with an acoustic or electric field, $\nu_a = \nu_i$, whereas the distributions remained unchanged when off-resonance (Fig. 4, solid line).

The parameters of the resonant events were faithfully reproduced at each unfolding-refolding cycle, and resonance was also found when using harmonics of ν_i (Fig. 3, E and J). Unfolding and refolding cycles performed on the same protein, with or without external field, always exhibited one and only one of the characteristic oscillation frequencies ν_i, independent of the frequency in the previous cycle and of the nature of the applied electric or acoustic fields (Fig. 3, C and D). Furthermore, the applied fields did not change the GFP spectroscopic properties (lifetime, brightness, and emission wavelength) (*29*), nor its denaturation kinetics parameters (*31*).

As expected, the resonant effects depended on both the applied frequency ν_a and the field strength, whereas the application of an off-resonance field (e.g., at ν_a = 250 ± 100 Hz) did not produce measurable effects (Fig. 3G) unless the field frequency bandwidth was broadened (ν_a = 250 ± 180 Hz). In this case, the first part of the oscillation interval became more regular, with $\Delta\tau_0$ increasing from 42 ms to 53 ms (Fig. 3I). The duration of the oscillation under ac electric and acoustic fields reached its top value at resonance, 80 ± 4 ms, only when ν_i fell within the applied field frequency range. When ν_a moved away from resonance, $\Delta\tau_0$ dropped from the resonant to the unperturbed value (fig. S7A) with a Lorentzian trend around the resonant frequency. $\Delta\tau_0$ is also a function of the field strength and can be described by a sigmoid function (fig. S7B). In particular, at an electric field

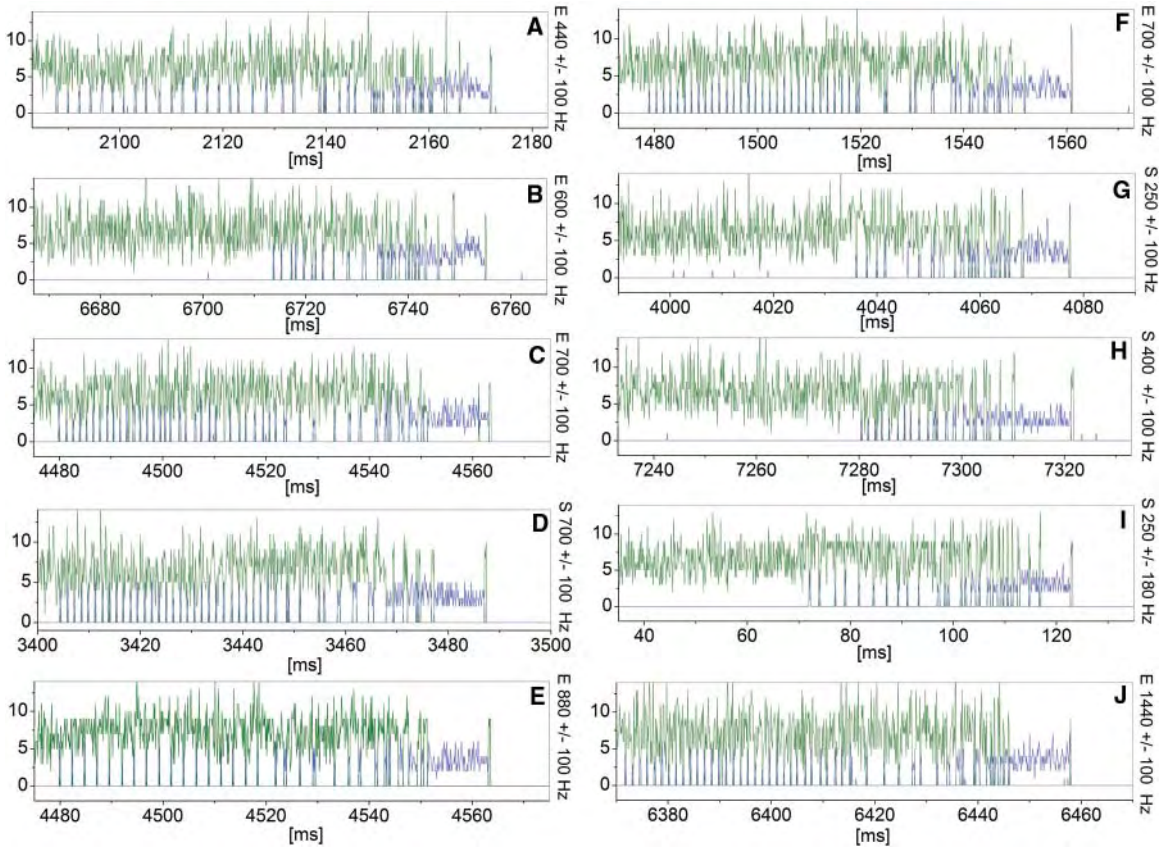


Fig. 3. Fluorescence pre-unfolding traces of two GFPs out of 102 proteins at 4.45 M GuHCl when electric or acoustic fields are applied during five consecutive unfolding-refolding cycles. The nature of the modulated perturbation (electric, E, or acoustic, S) and its frequency are indicated at the right of each plot. The green trace refers to the A channel and the blue to the N channel. (**A** to **E**) The first GFP. Cycles (A), (C), (D), and (E) show resonance either at ν_1 or ν_2. Cycle (B) does not show resonance behavior. (**F** to **J**) The second GFP. Cycles (F) and (J) show resonance at ν_2. Cycles (G) to (I) show no clear evidence of resonance behavior.

strength of ~0.9 V/cm, $\Delta\tau_0$ was found to lie halfway between the unperturbed and the sharper resonance value. The same relation pertains to $\Delta\tau_0$ versus acoustic field strength.

We note that the later part of the oscillating pre-unfolding fluorescence trace does not seem considerably perturbed by the applied fields. The data with and without applied fields are statistically equivalent in this regime.

The switching events among different conformational states of the folded GFP, as revealed by single-molecule fluorescence signals, were found to be randomly distributed in time, which suggests that thermal noise is the main cause of fluctuations (Fig. 1, E and G) (*3*, *14*). Previous observations of fluctuations in proteins and RNAs have been ascribed to the coupling between chemical kinetics and conformational transitions (*41*, *42*). This assumption has been the basis of some theoretical work (*15*, *17*) that implicates fast (picosecond) intramolecular fluctuations (*43*), coupled to chemical transitions, as indirect sources of slow (millisecond to second) fluctuations. Our own results show that (i) the random fluorescence fluctuations, recorded much earlier than the unfolding event, become nearly regular oscillations at one of the discovered GFP low frequencies, tens of milliseconds before the event; and (ii) marked resonant effects are induced by applied oscillating fields.

The regular fluorescence oscillations (~0.4 to 1 kHz) occur only in a narrow time window close to the instant of unfolding (Fig. 2D), with a transition to more chaotic behavior in the last milliseconds. In these late stages of the fluorescence emission, the barriers of the energy landscape are expected to become shallower than the energy separation of the vibrational levels, and therefore the electronic states can more strongly couple with some vibrational normal modes. At present, the complex nature of proteins' energy states obscures the details of the electronic and vibrational levels; consequently, a quantitative estimate of the vibronic coupling does not seem possible.

The involvement of a few characteristic fluorescence oscillation frequencies indicates that different chemical/conformational substates play a relevant role in the proximity of protein unfolding. In particular, the first part of the oscillation pattern, which displays regular alternations between the protein substates, suggests that a structured molecule with the A and N potential energy wells separated by a barrier (although in the process of collapsing because of the action of the denaturant) still exists, and hence periodic oscillations resonating with the applied fields can be expected. On the other hand, the irregular oscillations/fluctuations in the second part of the pattern appear related to the disappearance of the barrier structure under the final action of the denaturant. This view appears to be consistent with the observation that the final segment of the oscillations is unaffected by the applied fields.

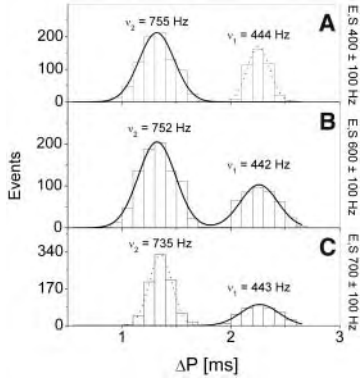


Fig. 4. ΔP distribution of the first part of the pre-unfolding fluorescence trace of 102 single GFPmut2 proteins at 4.45 M GuHCl under alternating electric, E, or acoustic field, S. (**A**) E and S field frequencies at 400 ± 100 Hz; (**B**) E and S fields at 600 ± 100 Hz; (**C**) E and S fields at 700 ± 100 Hz. The solid and dotted lines are Gaussian fits to the frequency distributions. Dotted lines correspond to resonant response of the proteins. The standard deviation of the distribution at resonance is approximately one-half of that off-resonance.

New experiments should accompany additional theoretical efforts in the direction of a deeper understanding of the coupling between chemical kinetics and conformational transitions in single molecules. Recent efforts to understand biological noise in cells, in which molecular species are modeled as discrete

entities through the use of master equations (*44*), might be adapted to the specific problems raised by the experiments presented here.

References and Notes

1. H. Frauenfelder, F. Parak, R. D. Young, *Annu. Rev. Biophys. Biophys. Chem.* **17**, 451 (1988).
2. H. Frauenfelder, S. G. Sligar, P. G. Wolynes, *Science* **254**, 1598 (1991).
3. L. Edman, R. Rigler, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8266 (2000).
4. Z. Xie *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 534 (2004).
5. D. S. Talaga *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13021 (2000).
6. A. A. Deniz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5179 (2000).
7. J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco Jr., C. Bustamante, *Science* **292**, 733 (2001).
8. W. E. Moerner, D. P. Fromm, *Rev. Sci. Instrum.* **74**, 3597 (2003).
9. A. Zumbusch, G. Jung, *Single Mol.* **1**, 261 (2000).
10. X. S. Xie, J. K. Trautman, *Annu. Rev. Phys. Chem.* **49**, 441 (1998).
11. T. Funatsu, Y. Harada, M. Tokunaga, K. Saito, T. Yanagida, *Nature* **374**, 555 (1995).
12. H. P. Lu, L. Xun, X. S. Xie, *Science* **282**, 1877 (1998).
13. T. Ha *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 893 (1999).
14. E. A. Lipman, B. Schuler, O. Bakajin, W. A. Eaton, *Science* **301**, 1233 (2003).
15. M. O. Vlad, F. Moran, F. W. Schneider, J. Ross, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12548 (2002).
16. R. Y. Tsien, *Annu. Rev. Biochem.* **67**, 509 (1998).
17. C. Scharnagl, R. Raupp-Kossmann, S. F. Fischer, *Biophys. J.* **77**, 1839 (1999).
18. A. B. Cubitt *et al.*, *Trends Biochem. Sci.* **20**, 448 (1995).
19. R. Heim, A. B. Cubitt, R. Y. Tsien, *Nature* **373**, 663 (1995).
20. T. M. Creemers, A. J. Lock, V. Subramaniam, T. M. Jovin, S. Volker, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2974 (2000).
21. W. Weber, V. Helms, J. A. McCammon, P. W. Langhoff, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6177 (1999).
22. M. Chattoraj, B. A. King, G. U. Bublitz, S. G. Boxer, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8362 (1996).
23. M. Zimmer, *Chem. Rev.* **102**, 759 (2002).
24. R. M. Dickson, A. B. Cubitt, R. Y. Tsien, W. E. Moerner, *Nature* **388**, 355 (1997).
25. S. Abbruzzetti *et al.*, *J. Am. Chem. Soc.* **127**, 626 (2005).
26. M. F. Garcia-Parajo, G. M. Segers-Nolten, J. A. Veerman, J. Greve, N. F. van Hulst, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7237 (2000).
27. U. Haupts, S. Maiti, P. Schwille, P. W. W. Webb, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13573 (1998).
28. B. P. Cormack, R. H. Valdivia, S. Falkow, *Gene* **173**, 33 (1996).
29. F. Cannone, M. Caccia, S. Bologna, A. Diaspro, G. Chirico, *Microsc. Res. Tech.* **65**, 186 (2004).
30. See supporting data on *Science* Online.
31. B. Campanini *et al.*, *Protein Sci.* **14**, 1125 (2005).
32. G. Chirico *et al.*, *Protein Sci.* **11**, 1152 (2002).
33. S. Bettati, B. Pioselli, B. Campanini, C. Viappiani, A. Mozzarelli, in *Encyclopedia of Nanoscience and Nanotechnology*, H. S. Nalwa, Ed. (American Scientific, Stevenson Ranch, CA, 2004), pp. 81–103.
34. J. D. Badjic, N. M. Kostic, *Chem. Mater.* **11**, 3671 (1999).
35. D. K. Eggers, J. S. Valentine, *Protein Sci.* **10**, 250 (2001).
36. D. K. Eggers, J. S. Valentine, *J. Mol. Biol.* **314**, 911 (2001).
37. C. Viappiani *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 14414 (2004).
38. C. Y. Shen, N. M. Kostic, *J. Am. Chem. Soc.* **119**, 1304 (1997).
39. Δt_{on} is the duration of a fluorescence burst; Δt_{off} is the time between two consecutive fluorescence bursts; ΔP is the time between the onset of two consecutive fluorescence bursts (*30*). For N-A switching events only, neutral channel Δt_{on} corresponds to anionic channel Δt_{off} and vice versa.
40. D. Silvano, S. Krol, A. Diaspro, O. Cavalleri, A. Gliozzi, *Microsc. Res. Tech.* **59**, 536 (2002).
41. L. Edman, U. Mets, U. R. Rigler, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6710 (1996).
42. M. Lim, P. Hamm, R. M. Hochstrasser, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15315 (1998).
43. D. ben-Avraham, *Phys. Rev. B* **47**, 14559 (1993).
44. C. V. Rao, D. M. Wolf, A. P. Arkin, *Nature* **420**, 231 (2002).

45. Supported by INFM, Fondo Integrativo per la Ricerca di Base, and Cofin Ministero dell'Istruzione Università e Ricerca. We thank A. Mozzarelli, S. Bettati, and B. Campanini for providing the GFPmut2 and the silica gel, and M. Collini for useful suggestions.

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5737/1096/DC1
Materials and Methods
Figs. S1 to S7
References

18 May 2005; accepted 11 July 2005
10.1126/science.1115001

Protein Expression System

TripleExpress is a rapid, flexible, and nonlytic insect-based protein expression system that avoids the manipulations and technical constraints of baculovirus systems. The patented triple combination of a powerful expression cassette (engineered from the silk moth cytoplasmic actin A3 gene, one of the strongest constitutive cellular promoters), an enhancer element, and a transcriptional activator enhances the production of recombinant proteins up to 5000-fold. With Triple Express, the quality of expressed proteins is enhanced through both reduced protein degradation and consistent posttranslational modifications. The system is amenable to both transient and stable production, and can be used for both small-scale and large-scale expression of membrane, cytoplasmic, and nuclear proteins.

CytoStore For information 888-863-2254 www.cytostore.com

Automated Image Analyzers

A new range of GeneGnome automated image analyzers is designed to make imaging chemiluminescent samples quick and simple. The GeneGnome comes complete with a 16-bit cooled camera inside a compact, light-tight cabinet and a new white light source for the added flexibility of imaging colored markers. The camera is fully integrated to a processor and flat screen monitor for easy sample viewing and utilizes the multiplexing capabilities of the system for fast and accurate molecular weight analysis. The GeneGnome is available in several versions, capable of imaging samples up to a maximum of 17 cm × 15 cm, making it simpler for users to analyze larger blots. The system is quick to set up. Its large fixed aperture lens means there is no need to spend time adjusting the camera so samples can be loaded into the system's automatic slide-out drawer; their images can be instantly captured with just one click.

Syngene For information 800-686-4407 www.syngene.com

High-Throughput Electrophoresis

The HTS Next Gel Kit is suitable for high-throughput electrophoresis of proteins. The kit contains a unique blend of agarose and Next Gel Running Buffer and can provide extended resolution of sodium dodecyl sulfate denatured proteins using standard horizontal gel apparatus. Various configurations of the horizontal system allow for the analysis of 20 to 200 samples in 1 to 2 hours.

Amresco For information 800-448-4442 www.amresco-inc.com

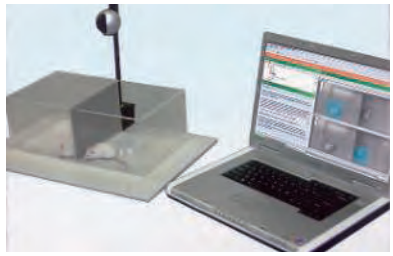
Tracking Software

Inventory of valuable biological samples and specimens is critical for basic research laboratories and biotechnology businesses. Cryotrack offers the ability to track large numbers and types of samples (DNA, RNA, plasmids, clones, proteins, peptides, probes, antibodies, enzymes, tissues, cell lines, and more). Users benefit from numerous features, including snapshots, graphical view, import and export of data, edit history, creation of replicates, frequently used fields selection, spreadsheets, extensive search options, barcode compatibility, ability to track and associate notes, and printing labels. The software is customizable to fit users' needs. This personal computer-based system is a robust client/server application that allows multiple users to access the same data while protecting the integrity of data by password security.

CryoTrack For information 650-941-6774 www.cryotrack.com

Maze Tracking System

ANY-maze is an advanced behavioral video tracking system that makes it possible for researchers to automate and track animals in virtually any maze or cage system using a USB digital camera and conventional notebook computer. Researchers can also choose to create more advanced systems with multiple apparatuses and sophisticated cameras. ANY-maze is capable of tracking animals in 16 different pieces of apparatus simultaneously. It includes more than 100 standard measures covering such things as distance, speed, and zone entries, and provides graphic and statistical analyses of results immediately after testing.



Stoelting For information 630-860-9700 www.stoeltingco.com/physio

cDNA Normalization

Complementary DNA (cDNA) normalization results in equalization of the abundance of different transcripts and an increase in the number of previously undetected genes in the cDNA library. This increases the efficiency of transcriptome analysis and functional screening of cDNA libraries. The Trimmer-Direct kit is a novel, effective way to normalize cDNA enriched with full-length sequences. The kit is based on technology that makes use of the unique properties of Kamchatka crab duplex-specific nuclease.

Evrogen For information +7 (095) 336 63 88 www.evrogen.com

Kit for Difficult Proteins

The PAGEprep Advance Kit improves sodium dodecyl sulfate analysis of difficult-to-analyze proteins samples by eliminating band distortion. The kit makes use of a unique spin-cup format and an optimized procedure that can achieve 75% to 85% protein recovery for most protein samples (1 to 70 μg) in less than 20 min. The kit is especially useful for inclusion bodies solubilized in guanidine-HCl; samples containing low-pH buffers, thiocynate, or urea; proteins precipitated in ammonium sulfate; and protein separation from chaotropic agents, detergents, DNA, RNA, lipids, and salts.

Pierce For information 800-874-3723 www.piercenet.com

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and government organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier by visiting www.science.labvelocity.com on the Web, where you can request that the information be sent to you by e-mail, fax, mail, or telephone.

Published by AAAS