nature

YOUNGER FOR LONGER

How dietary restriction throws the longevity switch

OCEAN MIXING
Hurricanes making waves

ANIMAL PERSONALITIES
Evolution of individuality

MATHEMATICAL ABILITY
The kids are alright

NATURE.COM
Research highlights

nature

www.nature.com/nature

Vol 447 | Issue no. 7144 | 31 May 2007

No more hot air

The leaders meeting at this year's G8 summit must grasp the opportunity to assert themselves and commit to real action on climate change.

When the world's most powerful political leaders convene at the G8 summit next week in the German spa town of Heiligendamm, they will bring with them pre-prepared communiqués on most of the topics to be discussed, from the financial risks of globalization to the need for development aid in Africa.

But the eight heads of states will also carry with them responsibility for most of the world's annual greenhouse-gas emissions. Angela Merkel, the German Chancellor, who hosts this G8, wants the leaders at Heiligendamm to agree a concrete plan on how to substantially lighten this load in the next couple of decades.

The Bush administration, however, seems once again to be working to foil any meaningful progress by the G8 on climate issues. Merkel should learn lessons from what happened to UK prime minister Tony Blair when he sought to pursue the same agenda at the G8 at Gleneagles, Scotland, two years ago: by accommodating US resistance and talking compromise, he achieved precisely nothing.

This time, Merkel should hold her ground, refuse to include inadequate climate-change language in the final communiqué and, if necessary, dismiss G8 protocol and break publicly on the issue with Bush and any allies he can muster. She should be encouraged in such a stance by the presence of US House speaker Nancy Pelosi (Democrat, California), who is visiting Europe this week. Pelosi's trip, in effect, aims to remind both Europeans and her supporters at home that the Bush administration no longer speaks for America on the climate-change issue.

The G8 leaders are uniquely placed to confront the issue of global warming. Negotiations at the United Nations' upcoming climate summit in Bali will be led by environment ministers, diplomats and subordinate government delegations. They are doomed to failure in the absence of a clear and unambiguous political mandate from above.

The G8 summit can best achieve that by stating unequivocally that the negotiations in Bali must achieve a robust and effective follow-up to the 1997 Kyoto Protocol, which required countries to reduce their greenhouse-gas emissions by an average 5% relative to 1990 levels in the 2008–12 commitment period.

"The G8 leaders are uniquely placed to confront the issue of global warming."

Such a follow-up agreement needs to include the active participation of the United States and timelines for the involvement of India and China. It will probably involve fresh, mandatory caps on emissions and an expanded cap-and-trade scheme modelled on Europe's emerging carbon market, modified to incorporate tax-based incentives to reduce emissions. Kyoto may have been a flawed agreement but there is no going back on the concept of an international treaty, led by the developed countries but involving developing ones too, as a central component of a global strategy to curtail emissions.

Participants in the G8 summits have built these gatherings up, over many years and in the face of considerable public scepticism, as the very pinnacle of global democratic leadership in the developed world. Failure to lead on the climate issue next week can suggest only that the scepticism was justified. ■

Time for a medical

The pharmaceutical industry is struggling to adapt to a harsher political environment.

Last week was an acutely difficult one for the pharmaceutical industry. On 20 May, Pfizer, the world's biggest drug company, announced the departure of both its research director, John LaMattina, and its chief financial officer. The following day, a study in the *New England Journal of Medicine* cast doubt on the safety of GlaxoSmithKline's blockbuster diabetes drug Avandia (see page 512). And that same day, Amgen received an unwelcome subpoena from the New York State attorney general, apparently related to questions about its marketing activities that have so far this year knocked one-fifth off the immense market capitalization of one of the world's top two biotechnology companies.

If these were just three isolated incidents, they might be of no great concern to the thousands of researchers who work for major drug manufacturers around the world. But the nature of the events themselves, and the way they've been received in the industry's largest and most lucrative market — the United States — carry important warning signs for the industry as a whole.

The barrage of bad news comes as the industry is trying to grapple with a new and more problematic environment for its business in the United States. Trouble over regulation, in particular, has been brewing for a few years now. But it is the change in control in Congress after last November's election that challenges the industry most directly.

"There is a sense that the heady growth that the pharmaceutical sector enjoyed in the 1990s is not going to be revisited."

Action is already pending in Congress to strengthen the regulator, the Food and Drug Administration, to make provision for the approval of biogenerics and, perhaps most ominously of all, to shift the entire patent regime in a direction that will please the information-technology industry (which likes patent sharing) at the expense of the drug industry (which does not). On top of that, there is the prospect of a new push towards general healthcare reform after the next presidential election, and the associated prospect of

price controls, akin to those in force in most nations outside the United States.

In heading off these challenges, the industry has several ingrained advantages. It is widely and rightly seen as an important engine for innovation, and therefore for US economic strength, and it can muster a formidable coalition of allies, ranging from doctors and patient groups to medical schools and the cities that host them.

But it also has some disadvantages — most notably its track record of aligning itself more closely than any other major industry, save oil, with the Republican Party. According to the *Wall Street Journal*, 69% of the industry's political contributions in last year's midterm election went to Republican candidates. It was seen as a necessary bet, given the industry's fear of tighter regulation, but it turns out not to have been a prudent one.

That, coupled with continuing public discontent about healthcare costs in the United States, has put the industry firmly on the back foot this spring. It is underperforming in the stock market, where there is a sense that the heady growth that the pharmaceutical sector enjoyed in the 1990s is not going to be revisited.

The industry's response to all of this has yet to take shape. It is trying to align its public image more closely with that of its biotechnology component (innovative, science-based, responsive to patient needs) and it has tried to introduce some self-regulation, in areas such as direct-to-consumer advertising. At the same time, the changes at Pfizer and similar developments elsewhere point to at least the possibility of a major consolidation of research and development activity to fit straightened circumstances (see *Nature* **445,** 13; 2007).

In terms of science and innovation, the pharmaceutical industry's best days ought to be ahead of it. The sequencing of the human genome and parallel developments in cell biology and immunology should greatly increase the potential for developing effective therapeutics, including ones matched to individuals' genetic make-ups. But the industry has some tricky terrain to navigate before these days arrive — and it is by no means clear that today's big-name companies will be around to enjoy them. ■

The safety catch

The United States' domestic security agency has yet to make best use of science and technology.

The US Department of Homeland Security (DHS) came into being in less than auspicious circumstances, as President Bush and Congress each sought to appear responsive to the terrorist attacks of 11 September 2001 by establishing a government department charged with securing the homeland.

When the department finally emerged in the spring of 2003, its composition reflected a series of unsatisfactory political compromises. The crown jewels of the state's security apparatus — the FBI and the CIA — were left alone, and the DHS emerged as a hodgepodge of the rest, ranging from the coast guard to the president's own security detail.

In an effort to ensure that the new department would channel the best of science and technology into its homeland-security mission, it was equipped with its very own science and technology directorate. It was hoped that the directorate would lend an up-to-date, cutting-edge flavour to the new department, reflecting the hope, widespread in the United States, that science and technology could be effectively harnessed to fend off terrorist attacks.

But relatively little of the counterterrorism work done by the department involves high technology. Although opportunities exist to use technology to improve performance at the margins, much of the work is about the efficient application of simple techniques. Patrolling the borders requires little more than a pick-up truck and a pair of binoculars; managing immigration paperwork plays to the skills of adept clerical staff, not turtlenecked hackers; and patrolling a coastline can be done as well in a 1950s-era cutter as it can in a hovercraft.

Add to the mix the fact that the new department's 'gang of seven' independent agencies often don't get along, and it is no surprise that the directorate has struggled to establish direction. As reported on page 516, it has always lacked a clear mission and, in the absence of one, has instead undertaken a variety of odd jobs for these component agencies, such as monitoring cities for biological agents and developing an anti-missile system for commercial aircraft. It has never been effectively managed, delivering reports to Congress late or not at all, and failing to account adequately for its spending.

Jay Cohen, who was appointed as undersecretary for science and technology at the DHS last August, is charged with reversing this track record. Cohen's vision for the directorate is a pragmatic one, concentrating on the development or acquisition of the technologies that the department needs to do its job. He accepts nonetheless that basic scientific research has a role in the directorate.

"The Department of Homeland Security is deeply fragmented — some would say dysfunctional."

Cohen's first nine months on the job have shown some promise. He has succeeded in bringing to the table representatives from the gang of seven to talk about what technologies they could actually use. Their needs are fairly basic — the immigration service, for example, would like better database software to manage its files — but they provide a framework that can lend some badly needed direction to the staff and grantees of the science and technology directorate.

It remains to be seen whether the directorate can deliver. The DHS is deeply fragmented, some would say dysfunctional, as its notoriously weak response to 2005's Hurricane Katrina demonstrated. And the science and technology directorate's performance has so failed to impress Congress that its budget was slashed by a quarter, to $848 million, this year. Morale has been low, and the directorate has failed, in its first years of existence, to forge strong staff leadership or a clear identity.

Cohen seems to be full of ideas and verve — but the fact remains that he is running a small appendage on an unwieldy department. As long as the DHS itself remains adrift, it is hard to envisage how its science and technology directorate can excel. ■

RESEARCH HIGHLIGHTS

PARTICLE PHYSICS

Reality check

Phys. Rev. Lett. **98,** 201801 (2007)

Are axions for real? These hypothetical particles, proposed originally in theories of the nuclear strong force, are now candidates for dark matter. Malcolm Fairbairn of Stockholm University in Sweden and his co-workers have put forward a new way to look for them.

Theory predicts that, in a strong magnetic field, photons convert to axions and back again. This conversion could happen in the Sun, the authors suggest, making it transparent to some forms of radiation. If axions with certain properties exist in the Sun, about 2% of high-energy γ-rays falling on the Sun's surface should travel right through.

Telescopes might detect this effect when the Sun passes in front of a distant γ-ray source. Current observations are not conclusive, the team says, but a more sensitive γ-ray satellite currently under construction — called GLAST — could resolve the issue.

ATMOSPHERIC SCIENCE

Cloudy skies ahead

Geophys. Res. Lett. **34,** L09811 (2007)

Early results from NASA's CloudSat (pictured below) are helping to clarify details of how clouds are distributed around the planet and vertically within the atmosphere. Such information is important for building accurate global climate models.

The first three months' worth of data collected after the probe's launch in April 2006 show that clouds over the tropical oceans fall into two main categories, report John Haynes and Graeme Stephens of Colorado State University in Fort Collins. These comprise clouds whose tops are about 2 kilometres above the Earth's surface, and those that top out at around 12 kilometres. In addition, a layer of cloud in between those two heights was found to be common over the western Pacific ocean.

NASA/JPL

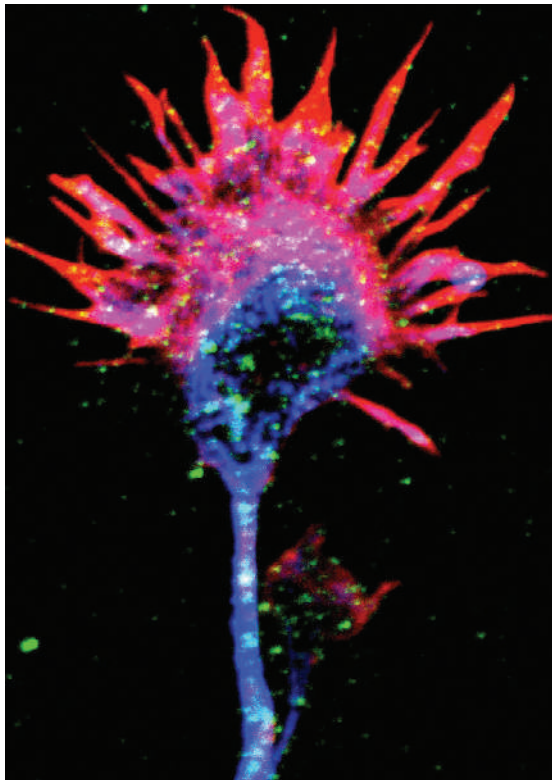
Trippy tips

Science **316,** 1212–1216 (2007)

Cannabinoids may be best known as the active ingredient in marijuana, but their endogenous forms — made naturally in the brain — have a different role. New work suggests they help to guide developing nerve fibres.

Tibor Harkany of the Karolinska Institute in Stockholm, Sweden, and his team tracked the cannabinoid receptor CB_1R in neurons in cell culture (pictured). They found that CB_1Rs (green) were recruited to the growing tip (pink) of the fibre (blue). Activation of CB_1Rs repelled the nerve fibres, encouraging them to grow away from the direction of the cannabinoid signal.

Deleting CB_1Rs in specific neurons in live mice changed the pattern of connections between axons. The authors speculate that in pregnant women, taking marijuana may interfere with the fetus' cannabinoid signalling, affecting neuronal development.



SCIENCE

Clouds that produced rain — 18% of those detected — were markedly thicker than clouds that didn't.

PHYSICS

Best served chilled

Phys. Rev. Lett. 98, 200801 (2007)

A device that can map tiny magnetic fields more precisely than existing techniques has been created from a quantum form of matter known as a Bose–Einstein condensate.

Physicists at the University of California, Berkeley, made a Bose–Einstein condensate by chilling a gas of rubidium atoms, held in place by laser light, to close to absolute zero. By imaging its response to magnetic fields, they could measure and map low-frequency fields with much greater sensitivity than today's best magnetometers. The device could be used to study inert materials or the magnetic fields produced by living tissues.

DEVELOPMENTAL BIOLOGY

Nose left from right

Cell **129,** 787–799 (2007)

The worm *Caenorhabditis elegans* has 'gap junctions' to thank for the asymmetry of its olfactory system, researchers report.

Cornelia Bargmann from the Rockefeller University in New York and her colleagues have found that more than 10% of developing neurons in the worm are temporarily connected into a network by gap junctions. These are channels that let small molecules and electrical signals pass between neurons.

Similar transient networks have been seen in embryos of other organisms, but their function was unclear. In the worm, the gap junctions help to organize the asymmetric olfactory system, which has different receptors for food-related smells in neurons in its left and right side. The transient connections also shape the arrangement of synapses in the final neural network.

METHODS

Sugar on the brain

Nature Chem. Biol. doi:10.1038/nchembio881 (2007)

A new technique can spot proteins that have been modified by the addition of the sugar β-*N*-acetyl-D-glucosamine (*O*-GlcNAc).

Linda Hsieh-Wilson of the California Institute of Technology in Pasadena and her colleagues have used the method to show that *O*-GlcNAc modification may be important in neuronal communication.

The technique combines labelling of the protein of interest and the sugar group to make the modified proteins quantifiable by mass spectrometry.

Hsieh-Wilson's group studied a set of

neuronal proteins. For several, the method could pin-point the position of the added sugar. They also found that some proteins showed more *O*-GlcNAc modifications if they were taken from an intact brain shortly after its neurons had been stimulated than if they were taken from unstimulated brain tissue.

ASTRONOMY

Universal dust-up

Astrophys. J. **661,** L9–L12 (2007)

The most distant γ-ray burst ever seen has cast its light on dust in the early Universe.

Giulia Stratta of the ASI Science Data Center in Frascati, Italy, and her colleagues studied the radiation from a 12.8-billion-year-old burst over three days. The emission was fainter at certain wavelengths than expected from the burst's initial brightness, suggesting that dust in the early Universe is different to that found in the Universe today.

Dust is produced in the dying explosions of stars, so the amount of dust present at this early time and its composition could provide clues about how the first stars formed.

CHEMISTRY

The simplest link

Science **316,** 1172–1175 (2007)

It is now possible to join certain ring-shaped molecules together without resorting to chemical tinkering to make them more reactive. This provides a simple way to perform 'cross-coupling' reactions, a type of reaction widely used in the drug industry.

David Stuart and Keith Fagnou from the University of Ottawa, Canada, used a palladium and copper catalyst system to build carbon–carbon bonds between benzene and two-ringed molecules known as indoles.

Cross-coupling reactions have previously required several steps, with the starting materials first being converted into more reactive analogues. In the new scheme, the catalyst activates a carbon–hydrogen bond on the indole, making it reactive enough to form a bond with one of the benzene's carbon atoms. There are no unwanted side products.

CANCER BIOLOGY

Stem cells fished out

Genes Dev. doi:10.1101/gad.1545007 (2007)

A model of a common and often fatal childhood cancer — embryonal rhabdomyosarcoma (ERMS) — may have helped researchers to identify the stem cells that mediate the disease.

Leonard Zon of the Children's Hospital Boston in Massachusetts and his colleagues induced ERMS in zebrafish by activating a signalling pathway that is mediated by the Ras protein. This pathway is commonly activated in human ERMS, the researchers found.

They identified genetic pathways that drive progression of the disease in both zebrafish and humans, and found that tumour development in the zebrafish depends on a population of cancer stem cells. These cells triggered tumour development when transplanted into healthy animals. The human counterparts may be 'activated satellite cells', found in muscle. Gene-expression studies showed these to have similar self-renewal mechanisms to the zebrafish cancer stem cells.

T. BLACKALL

NITROGEN CYCLE

Seabirds add ammonia

Geophys. Res. Lett. **34,** L10801 (2007)

Seabird colonies are the world's largest point sources of atmospheric ammonia, according to new calculations.

Trevor Blackall, now at King's College London, and his colleagues travelled to two Scottish islands — the Isle of May, home to a colony of Atlantic puffins, and Bass Rock (pictured above), which houses thousands of Northern gannets — to measure how much of the gas is released by bird droppings.

Globally, birds' ammonia emissions are outstripped by those from livestock, synthetic fertilizers and oceans. But the researchers estimate that, in the relatively pristine Southern Ocean below 45° S, penguins account for almost 20% of ammonia emissions. The largest colonies may produce up to 6,000 tonnes of ammonia per year, more than even the biggest poultry farms.

JOURNAL CLUB

Iwao Ohmine
Nagoya University, Japan

A theoretical chemist compares love to hydrogen bonds.

Water molecules assemble into ice "palm to palm", like Romeo and Juliet on their first encounter. Each molecule reaches out to four neighbours, forming hydrogen bonds that lock the molecules into a tetrahedral network. And like the love of Shakespeare's pair, water's hydrogen bonds are resilient. Ice contrives to keep its network, even in the tightest of spaces.

Researchers recently predicted that ice constrained by a carbon nanotube's wall will form either tubular structures or intricate arrangements of double- and quadruple-stranded helices, depending on temperature, pressure and nanotube diameter (J. Bai *et al. Proc. Natl Acad. Sci. USA* **103,** 19664–19667; 2006).

I have spent many years studying the structure and dynamics of water, but am still amazed by these luxuriant ice structures. Had computer simulations not shown how strenuously ice's network can adapt for its molecules to keep their four hands touching, we could hardly have imagined such structures would be possible.

Simulations have also predicted that confined ice can have two symmetrically different phases, which become deformed and indistinguishable when put under pressure (K. Koga *et al. Nature* **412,** 802–805; 2001). So we expect that one type of ice will easily transform into the other through collective motion of its hydrogen bonds.

My prediction is that confined liquid water, which has a disordered network of hydrogen bonds, will undergo similar structural rearrangements. Molecular mechanisms may cause large changes to the network structure of water trapped in proteins or at membrane surfaces, for example. These studies could therefore help us begin to understand another intimate relationship — the relationship between water and life.

Discuss this paper at **http://blogs.nature.com/nature/journalclub**

NEWS

Weighing up the evidence

A meta-analysis of clinical trials for the diabetes drug Avandia has hinted at possible cardiovascular risks; but how clear is the study, and how should meta-analyses be viewed against clinical trials?

The day after concerns about Avandia hit the news, the emergency room at New York City's Bellevue Hospital was even more hectic than usual. Patients with diabetes filed in and asked to discontinue their medication while the phone rang with calls from other Avandia users. Harried doctors rushed to handle the extra flow.

Cardiologist James Underberg found the scene disturbing. Although he insists that he is a clinician, not a statistician, Underberg could see that the analysis that suggested Avandia might be causing heart attacks was plagued by limitations — the authors of the paper said so themselves. But he could now be sued the next time he writes a prescription for Avandia, he says, "all on the basis of one article that may or may not have significant flaws in the analysis".

J. SULLIVAN/GETTY

An analysis of Avandia trials has highlighted potential side effects.

The article, published last week (S. E. Nissen and K. Wolski *N. Engl. J. Med.* doi:10.1056/NEJMoa072761; 2007), is a meta-analysis of 42 clinical trials of rosiglitazone, a drug used to lower blood sugar in patients with diabetes and marketed under the name Avandia by London-based pharmaceutical giant GlaxoSmithKline (GSK). The authors, Steven Nissen and Kathy Wolski of the Cleveland Clinic in Ohio, found that the chance of having a heart attack was 43% higher in patients who were taking the drug than in those who weren't, although precisely why is unclear (see 'How Avandia works').

These data do not differ strikingly from others that have come before; in Europe, Avandia comes with a warning label that the drug increases the risk of heart attack by 31%, on the basis of a previous meta-analysis, by GSK. But the new paper has caught a lot of attention. *The Lancet*, which had earlier published the promising results of one of GSK's clinical trials of Avandia, responded with an editorial warning against alarmist responses to Nissen and Wolski's paper and its tone of "urgency" (*Lancet* doi:10.1016/S0140-6736(07)60787-9; 2007). But by then, GSK's stock price had plummeted and congressional hearings had been scheduled to review the handling of the drug by the US Food and Drug Administration (FDA).

> "If I saw a meta-analysis done by GSK on Avandia, I don't think I would believe it."

Meta-analyses, which crunch together the results of many clinical trials, are powerful statistical tools to uncover the side effects of drugs. They were key in bringing to light the increased chance of heart attack associated with the painkiller Vioxx (also highlighted by Nissen, see 'Man on a mission'), and the added risk of suicidal thinking and behaviour in children taking some antidepressants. Few doubt that meta-analyses have their uses, but researchers caution that they should be read in the right light and, if possible, made more robust.

The studies crunched together in this case were not designed to look for cardiovascular side effects, but rather to assess the drug's main activity. Reports of heart incidents may not have followed the same rigorous criteria in all studies, nor were they reviewed by an outside panel. And Nissen and Wolski had access only to the trials' summaries, and not to patient-specific data (GSK has posted such summaries online since an unrelated court case in 2004 prompted it to do so), so any confounding effects could not be taken into account. If the cardiovascular risks are only for a specific subset of patients, this meta-analysis won't be able to tell.

To make such studies more robust, more data are needed. John Ioannidis, an epidemiologist at the University of Ioannina School of Medicine in Greece who has examined numerous meta-analyses notes that journal articles reporting clinical trials tend to dedicate more space to listing the authors' names than to listing possible side effects associated with the drug. And studies that produce negative results are often not published at all, giving researchers an incomplete picture of how often trials fail. Ioannidis argues that use of complete data sets would strengthen the analysis and could weed out potential false alarms. "I think that companies would actually gain from such transparencies," he says.

He adds that meta-analyses should be done by independent groups. At the moment, they are frequently conducted as part of the drug approval process. But in many such cases, the analysis is being done by the pharmaceutical companies that manufacture the drug in question — something that most agree is not a good idea. "If I saw a meta-analysis done by GSK on Avandia, I don't think I would believe it,"

How Avandia works

How does a drug that lowers blood sugar levels cause heart attacks? Avandia (rosiglitazone) is a member of a troubled class of drugs called thiazolidinediones that activate a set of receptors in cell nuclei. These receptors — the peroxisome proliferator-activated receptors — then modulate the expression of multiple genes. The full physiological consequences of tweaking this expression are not known, and it comes as little surprise that the drugs have multiple effects. Another drug in this class was taken off the market because it caused liver damage, and attempts to develop other thiazolidinediones have encountered problems with other harmful side effects.

Rosiglitazone raises levels of low-density lipoprotein cholesterol — the 'bad cholesterol' that is associated with cardiovascular problems. The drug also slightly lowers haemoglobin concentrations, which can increase stress on the heart. Most doctors agree that the magnitude of these changes wouldn't be sufficient to boost the risk of heart attack by 43%, but as the absolute value of heart-attack risk calculated by Nissen's meta-analysis is in question, both of these mechanisms remain possible culprits. **H.L.**

PLASTICS FOR POSTERITY
How do you prevent valuable collectors' items from degrading?
www.nature.com/news

NATIONAL MUSEUM OF DENMARK

says Ioannidis. "Someone who is independent should do this."

But GSK doesn't think that crunching studies together is in itself the best way to find side-effects. "We don't think meta-analysis is the best way to address this," says GSK spokeswoman Alice Hunt. Clinical trials, she argues, are the only way to achieve a definitive answer. GSK is in the middle of a clinical trial of more than 4,000 patients that is aimed specifically at assessing the cardiovascular effects of Avandia (compared with more than 10,000 patients in Nissen's number crunch). *The Lancet* editorial also recommends waiting for these results — although the news of the new findings has reportedly led to volunteers dropping out of the trial.

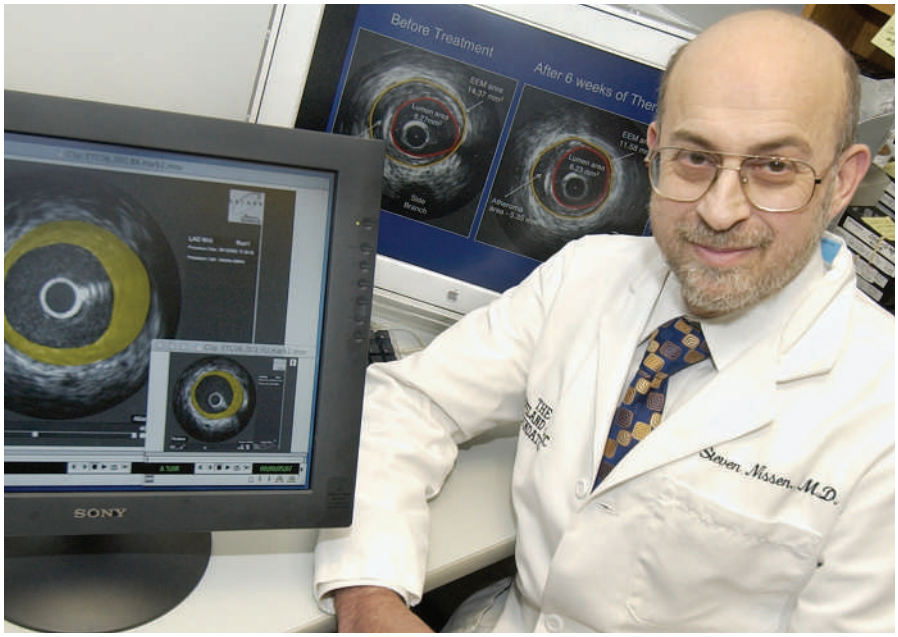
Preliminary results of the trial, scheduled to finish in 2008, have been shared with the FDA, says Hunt. "They could potentially be released, but that's not being looked at, at the moment," she adds.

The FDA will not as yet comment on the ongoing trial, to "preserve the study integrity", but it does have additional data on the drug. Without seeing the results for themselves, some doctors find it hard to know how to judge FDA assurances that the additional data it has do not point to a significant risk of heart attack. "They are unpublished, uncited data that are not available to the public. I've heard the same story before: 'We know there are weapons of mass destruction in Iraq. We can't tell you how we know, but we know'," says Nissen.

For some, even a preliminary analysis is reason enough to avoid the drug, if only because there is an alternative medication — pioglitazone, marketed under the name Actos in the United States — that has not yet been associated with cardiovascular risks. Nissen's analysis was enough to convince Harlan Krumholz, a cardiologist at Yale University: "When you're talking about safety and you're talking about a drug for which there are alternatives, you have to ask yourself, who has the burden of proof here?"

Although Avandia's fate remains unclear, Nissen's study has served to raise the profile of meta-analyses. Roger Chou, a clinician at the Oregon Evidence-based Practice Center in Portland notes that at the moment, independent analyses such as this are done on an ad-hoc basis. "Right now, it's really just when somebody's interested in it, or when there's something that's making people concerned," he says. But there is a growing interest in independent groups monitoring drug safety in this way — and, importantly, signs of increased funding from organizations such as the US Agency for Healthcare Research and Quality to support meta-analyses. ■

Heidi Ledford



A. SANCETTA/AP

Man on a mission

It isn't every clinical scientist who in the space of a month sends a major drug company's stock plummeting, is invited to testify at Congress and is crowned one of the world's 100 most influential people by *Time* magazine. But then, they aren't Steven Nissen, the cardiologist who last week fingered the diabetes drug Avandia as carrying a possible risk of heart attacks (see opposite).

"He will persevere for what he believes is right."

This is not the first or most famous case for Nissen, 58, who is based at the Cleveland Clinic in Ohio. In 2001, while serving on a committee of external advisers to the US Food and Drug Administration (FDA), he raised concerns about the safety of a painkiller called Vioxx. Before then, he had made his name pioneering an ultrasound technique that allowed doctors to see fatty plaques of atherosclerosis (see *Nature Med.* 11, 700; 2005).

Since Vioxx, Nissen says that he has become more concerned about the scientific rigour of drug regulation, prompting him to dig into clinical-trial data that suggest hidden drug risks. "I don't go looking for these things, but they sure seem to find me," he says. Between running trials for several drug companies, he has flagged cardiovascular risks in drugs for everything from attention deficit hyperactivity disorder to heart failure itself, often costing companies millions in the process.

His fans praise him for refusing to pull his punches. "He will persevere for what he believes is right, regardless of the toes on which he may tread," says Peter Libby, chief of cardiology at Brigham and Women's Hospital in Boston, Massachusetts, who has worked with Nissen for twelve years.

But his methods have earned him enemies, too, many of whom contend that he is more interested in felling the next Goliath than in seeking scientific truth. "The caped crusader Nissen is at it again," groused a critic on the *Wall Street Journal's* health blog on 21 May. Bob Temple, director of medical policy at the FDA's drug review centre, has also taken a shot at Nissen. Temple told *RPM Report* that he was "sort of stunned" by Nissen's suggestions to Congress that senior FDA officials had overruled their underlings' safety concerns. "I can tell you, he didn't see that," Temple said.

Nissen certainly isn't always right. Speaking to journalists last October, he put Pfizer's cholesterol drug torcetrapib, for which he was overseeing a clinical trial, at the top of a list of hot stories that reporters should watch in 2007. Weeks later, Pfizer pulled the plug on the drug when it was found to raise death rates in late-stage trials (see *Nature* 444, 794–795; 2006).

"After the torcetrapib business, I thought I would lay low for a while," Nissen said last week. "Then I stumble across this [Avandia] problem and here we are again." ■

Meredith Wadman

Stormy opening to hurricane season

The US National Hurricane Center in Miami, Florida, always uses the 1 June start date of the Atlantic hurricane season to remind coastal residents to be prepared for storms. But the lead-up to this year's season has exposed internal bickering and raised questions about whether the centre has the resources it needs.

Just days before the federal government issued a dire prediction of an above-average hurricane season — similar to the 2006 prediction (see table) — the centre's director complained very publicly about the budget he had to track and forecast those storms. Bill Proenza, who came on board as centre director early this year, criticized his bosses for spending too much money on anniversary celebrations and not enough on replacing a key satellite — the Quick Scatterometer, or QuikSCAT. At the press briefing to announce hurricane predictions on 22 May, tensions ran high as Proenza shared the stage with the head of the National Oceanic and Atmospheric Administration (NOAA), retired vice-admiral Conrad Lautenbacher.

The spat reveals some long-simmering disputes between the various agencies that make up NOAA. The agency was cobbled together in 1970 from a number of existing groups (the oldest being the Coast and Geodetic Survey, which dates back to 1807 and is the basis for NOAA's '200th-anniversary' celebrations this year). As often with such unions, it hasn't always been a happy one.

Proenza and some other researchers complain that NOAA is trying to usurp the identities of the individual services that make it up. The agency has, for example, proposed changing the names of the National Hurricane Center and its parent, the National Weather Service, to the NOAA Hurricane Center and the NOAA Weather Service.

"Fund the weather service or die."

Disputes about service names may sound petty, but there are bigger underlying problems. Proenza charges that NOAA spent a disproportionate $4 million on anniversary celebrations aimed at outreach, and not enough on forecasting. NOAA's budget is roughly $4 billion; of that, $6 million goes to the hurricane centre. An agency spokesman disputes the outreach figure, saying it has amounted to only $1.5 million over two years, and Lautenbacher says the centre gets enough to do its job.

"They both need more money," says James Baker, who headed NOAA between 1993 and 2001. He defends outreach, pointing out that the agency is not as well known as NASA even though its mission is arguably more critical to people's lives. "I used to say, 'fund the weather service or die'," he says. "Trying to publicize and get the message out is good for both NOAA and the weather service."

NOAA's $4-billion budget suffers particularly in comparison to NASA's $16 billion. Whereas the space agency is tasked with building and flying research satellites for Earth observation, NOAA is the agency in charge of turning them into long-term operational reality. In recent years, a series of reports from the National Research Council has identified problems in making smooth transitions from one to the other without loss of data in between.

Of particular concern for hurricane forecasting is the QuikSCAT satellite, launched by NASA in 1999. An instrument known as SeaWinds aboard the satellite measures the direction and speed of winds by bouncing microwaves off the ocean surface and measuring them on their return to the satellite. The resulting data are used to improve hurricane forecasts as storms gather speed and change direction.

But the satellite has passed its nominal five-year life, and Proenza has made a particular point of arguing for a successor. If SeaWinds

TROUBLE AHEAD IN THE ATLANTIC?

	Storms predicted in May 2006	Actual 2006 hurricane season	Storms predicted in May 2007
Tropical storms	13–16	9†	13–17
Hurricanes	8–10	5	7–10
Major hurricanes*	4–6	2	3–5

SOURCE: NOAA

* Winds of 178 kilometres per hour or greater.

† Not including Hurricane Zeta, in January 2006, considered part of the 2005 season.

Academics strike back at spurious rankings

A call by a group of US colleges earlier this month to boycott the most influential university ranking in the United States has shone the spotlight on the problem of institutional rankings. Experts argue that these are based on dubious methodology and spurious data, yet they have huge influence. But help is at hand: European academics are putting some rigour into rankings by tackling the problem themselves.

On 5 May, Douglas Bennett, president of Earlham College in Richmond, Indiana, and 11 other college presidents asked colleagues to refuse to fill out surveys for the *U.S. News & World Report*. That survey of institutions, they argued, "implies a false precision and authority that is not warranted by the data they use". Another 17 colleges have since signed up.

"All current university rankings are flawed to some extent; most, fundamentally," says Alan Gilbert, president and vice-chancellor of the University of Manchester in Britain. "But rankings are here to stay, and it is therefore worth the time and effort to get them right."

The rankings in the *U.S. News & World Report* and those published by the British *Times Higher Education Supplement* (*THES*) depend heavily on surveys of thousands of experts — a system that some contest. A third popular ranking, by Jiao Tong University in Shanghai, China, is based on more quantitative measures, such as citations, numbers of Nobel prizewinners and publications in *Nature* and *Science*. But even these measures are not straightforward.

Thomson Scientific's ISI citation data are notoriously poor for use in rankings; names of institutions are spelled differently from one article to the next, and university affiliations are sometimes omitted altogether. After cleaning up ISI data on all UK papers for such effects, the Leeds-based consultancy Evidence Ltd, found the true number of papers from the University of Oxford, for example, to be 40% higher than listed by ISI, says director Jonathan Adams.

Researchers at Leiden University in the Netherlands have similarly recompiled the ISI database for 400 universities: half a million papers per year. Their system produces various rankings based on different indicators. One, for example, weights citations on the basis of their scientific field, so that a university that does well in a heavily cited field doesn't get an artificial extra boost (see table).

The German Center for Higher

R. MAYER/S. FLORIDA SUN-SENTINEL/MCT/NEWSCOM

Bill Proenza wants less partying and more money for hurricane-watching satellites.

failed, he says, the accuracy of three-day hurricane forecasts would drop by about 16%.

A second version of SeaWinds was launched aboard the Japanese ADEOS-II satellite in 2003, but failed just months later. Another replacement, NASA's Ocean Vector Winds mission, was cancelled early this decade.

In their place, Lautenbacher points to the joint military–civilian Coriolis satellite. It carries an instrument known as WindSat, which uses a different technique from SeaWinds for detecting wind speeds and direction — rather than bouncing microwaves off the sea surface, it detects emission of microwaves from the surface.

SeaWinds has better spatial resolution than WindSat and works better under weak winds and rainy conditions, says its principal investigator, Timothy Liu of NASA's Jet Propulsion Laboratory in Pasadena, California. But WindSat does some things that SeaWinds cannot — it can provide multiple sets of data simultaneously, counters its lead scientist, Peter Gaiser of the Naval Research Laboratory in Washington DC.

Nevertheless, QuikSCAT is at least five years past the end of its nominal lifespan, and Congress has begun to get worried. Several members of Congress are pressuring NASA and NOAA for more details on what they plan to do when QuikSCAT fails.

Any potential replacement would take time to put together. QuikSCAT itself was hastily pulled together and launched just two years after a previous instrument failed without warning, but only because key parts had already been built for other purposes.

In the meantime, tensions will continue to bubble, although Lautenbacher makes the conflict seem welcome. "You have to remember that Mr Proenza just took over as the head of the hurricane centre, and he is known for being a very strong and forceful advocate for his programmes," he says. "And that's one of the reasons why we love him." ■

Alexandra Witze

See also the News Feature on page 522.

EUROPEAN RANKINGS: ONE WAY OR ANOTHER

University	*THES*[1]	Jiao Tong University[2]	Leiden ranking: no. of publications[3]	Leiden preferred ranking: field-weighted average impact[4]
Cambridge	1	1	1	2
Oxford	2	2	3	1
ETH Zurich	6	5	21	3
Lausanne	31	79 (tied with 43 others)	89	4
Leiden	32	22	27	23

ETH, Swiss Federal Institute of Technology

[1] http://tinyurl.com/29w3c8; [2] http://tinyurl.com/ywhebn; [3] http://tinyurl.com/yozree; [4] http://tinyurl.com/2z35rb

Education Development (CHE) also offers rankings *à la carte* for almost 300 German, Austrian and Swiss universities — users can query its online database by individual discipline, number of publications or student outcomes, for example. With European Union support, the CHE is expanding the system to cover all Europe. Adams says that more complex systems such as these are better than oversimplified summary tables, which policy-makers tend to take at face value.

The US Commission on the Future of Higher Education is considering creating a similar public database, which would offer competition to the *U.S. News & World Report*. Bennett says he will work with such groups that use "professional standards of measurement". ■

Declan Butler

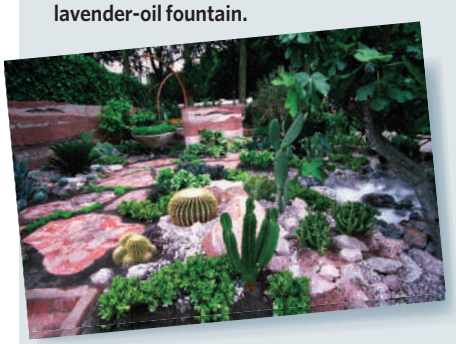
3 GOOD REASONS

Why science did well at this year's Chelsea Flower Show

1 Gold-medal-winning garden '600 Days with Bradstone' (pictured) was designed as a relaxation area — for a lonely astronaut on Mars.

2 A garden created in tribute to Linnaeus on the 300th anniversary of the naturalist's birth also scooped a gold.

3 And another gold medal went to an exhibit of plants with eco-industrial uses, including marigolds for paint, hemp for textiles — and, bizarrely, a lavender-oil fountain.



M. WALTER/TROIKA

SCORECARD

▲ **Everest's telecoms** British mountaineer Rod Baber has made the first mobile telephone call from the world's highest peak.

▼ **US television** Bloggers have panned CBS's forthcoming sitcom *The Big Bang Theory*. The highly original premise features two stereotypical physicists fumbling to impress their sexy new neighbour.

NUMBER CRUNCH

300 kilometres per hour is the top speed of Japan's new fleet of iconic bullet trains — the same top speed as the old ones.

19% is the energy saving promised by the new trains, which will run on far less electricity than their predecessors.

15 centimetres is the width of the arm-rests in first class, proving that going green doesn't always mean skimping on comfort.

Sources: Royal Horticultural Society, AFP, BBC, scienceblogs.com

SIDELINES

SPECIAL REPORT

Changing course

Science and technology have not always gone down well at the US Department of Homeland Security. **Geoff Brumfiel** reports on a retired Navy admiral trying to turn around the troubled research wing.

Thankfully, Jay Cohen's job did not begin with a bang. He assumed his post as undersecretary for science and technology at the US Department of Homeland Security (DHS), on 10 August 2006 — the day that British authorities reported a plot to smuggle liquid explosives aboard US-bound aircraft.

In the aftermath, the DHS rushed to restore international air travel and restrict liquids on flights. Any newcomer could have easily become lost in the shuffle, but Cohen is no shrinking violet. Within a day, he had organized a rapid-response team, tasked his scientists with understanding liquid explosives and briefed staffers and others on Capitol Hill. "By 11 August," he jokes, "my shyness had waned."

DHS

"I came into the job with my eyes wide open." — Jay Cohen

Shyness has little place at the science and technology (S&T) directorate, which has struggled for identity and purpose since its inception in 2003. In theory, it is the part of the department that comes up with new gizmos that allow border agents and customs officials to catch explosives or biological agents smuggled into the country. In practice, critics say, it has worked poorly with other divisions of the DHS and failed to account for its dollars. This year, a frustrated Congress slashed its budget by more than 25%, to $848 million (see bar chart).

The challenge for Cohen — a retired admiral who spent six years at the helm of the Office of Naval Research — is to restore a sense of purpose and productivity. Supporters say he is a welcome force for change. "I'm optimistic," says Congressman David Wu (Democrat, Oregon), who chairs the subcommittee overseeing DHS for the House Committee on Science and Technology. "Admiral Cohen's a breath of fresh air."

But others wonder whether Cohen will be able to turn around the directorate during his two-year appointment. "There is still no settled view on mission priorities," says Elizabeth Grossman, a former staffer on the committee now with the Washington DC-based lobbying firm Lewis-Burke Associates.

The DHS was officially born in January 2003, created by President George W. Bush as a response to the terrorist attacks of 11 September 2001 (see 'Counting the chemicals'). For the most part, Congress formed the $43-billion department — the third-largest in the federal government — by cobbling together existing agencies, such as the Coast Guard and the Secret Service.

An exception was the S&T directorate, which was created from scratch. In 18 pages of legislation, Congress laid out a plan for an S&T division to research future terrorist threats and support other directorates within the DHS. The scheme included plans for establishing university research centres and for collaborating with other agencies, notably the national labs run by the Department of Energy. "I think the directorate got off to a very good start," says Parney Albright, a former assistant secretary for science at the department from 2003 to 2005.

But as the directorate took over research efforts from some other large agencies — such as the customs service and the Transportation Security Administration, which oversees security at airports — it became ostracised from other divisions. It was slow in evaluating scientists' proposals and releasing funding (see *Nature* **424,** 986; 2003). And its honeymoon with Congress quickly soured, in part because it failed to produce documents that justified its spending.

Not everyone agrees on what led to these breakdowns. Albright attributes them to what he calls the lack of leadership from Cohen's predecessor, Charles McQueary (*Nature* **423,** 106; 2003). Other congressional staff and former departmental advisers point to the S&T division's newcomer status and high staff turnover.

The loss in faith was soon matched by an erosion of the directorate's power. In the spring of 2005, it lost control of research into the prevention of nuclear attacks, an area that was split off into a separate, roughly $300-million Domestic Nuclear Detection Office. "That took a big part of the portfolio away," says William Happer, a physicist at Princeton University, New Jersey, who served on a now-defunct scientific advisory committee to the department.

Rupture repair

Meanwhile, Congress and the Bush administration were saddling the directorate with other very specific projects, such as to investigate antimissile systems for commercial aircraft. In addition to diminishing the directorate's independence, the mandatory programmes gave it a piecemeal appearance. "It was potentially more ad hoc than one would wish," says Wu.

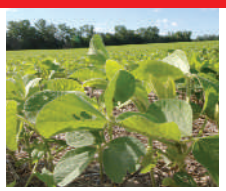
In March 2006, McQueary left to become chief of testing new weapons at the Pentagon. Later that spring, members of Congress released a damning appraisal of the S&T directorate, saying they were "extremely disappointed" by the agency's justifications of its programmes and "dismayed" by its bookkeeping.

Enter Cohen. "I came into the job with my eyes wide open," he says. His first priority was to realign basic research with the needs of homeland security's seven major agencies. He organized panels that included representatives from

Counting the chemicals

While the science directorate is getting its act together, the rest of the infant Department of Homeland Security (DHS) is busy irritating research chemists. Proposed new standards would force many universities to complete detailed inventories of the chemicals in their laboratories, and may eventually require 'site security plans' involving checks on employees and perimeter security.

Several research and education groups have already objected, saying the proposal is unsuitable for academic laboratories. The regulations require that any institution with certain quantities of specific chemicals complete a top-to-bottom inventory of chemical holdings. The American Chemical Society notes that some of the chemicals that would trigger the inventory, even in minute quantities — such as triethanolamine, carbon monoxide and hydrogen sulphide — are ubiquitous in research labs. The groups further argue that chemicals at universities are typically spread out in small quantities over a number of labs, making them less of a security threat but a bigger headache to count.

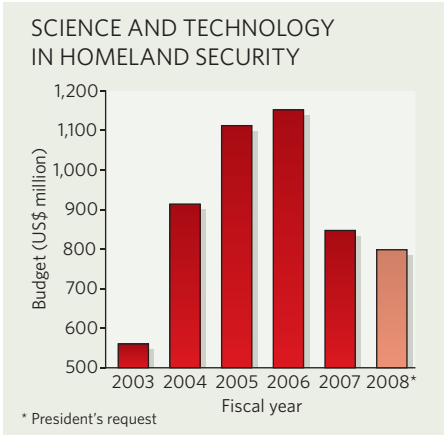


'NEXT GENERATION' OF TRANSGENIC CROPS
New strains provide more herbicide options.
www.nature.com/news

B. HAMPTON/UNIV. NEBRASKA

L SKOOGFORS/FEMA PHOTO

Jay Cohen's science wing plans to support homeland security missions, such as emergency management.



SOURCE: DHS

the 'gang of seven' and asked them to look at specific technical problems. He also reorganized the directorate, replacing the mission-oriented offices — such as for antimissile technology — with six subject-oriented divisions, such as infrastructure protection. Finally, he realigned the Department of Energy labs and directorate-funded university centres so that their research had a clear relationship to agency needs. "There must be an output," Cohen says. "Otherwise we're a self-licking ice-cream cone."

At a meeting last week designed to bring together universities, labs and department members — the first of its kind — representatives from the seven divisions generally praised his approach. Marko Bourne, director of policy at the Federal Emergency Management Agency, says that Cohen's office has helped unify research. Bourne's team is to develop new geospatial information tools that can help them to better plan disaster response. "Our interactions have been much more robust," Bourne says. "This works a lot better for us."

Cohen's other mission has been to repair relations with Capitol Hill. Prior to his stint as head of Naval Research, the talkative admiral served as a congressional liaison for the Navy and is well-known for his ability to work with Congress. "The one thing I'll say about Cohen — the guy's accessible," says one senior congressional staffer. "When you're in a meeting with him, you're in a meeting with him."

With his realignment complete and relations with Congress on the mend, Cohen says he now hopes to begin producing real results for the rest of the DHS. He wants to invest heavily in research on the psychology and sociology of terrorism. He also says that he would like to develop further technologies to detect improvised explosive devices and advance the use of composite materials in Coast Guard ships — two programmes he promoted heavily at the Office of Naval Research.

Cohen also plans on increasing basic research at the DHS. Through reprogramming, he has already doubled basic research from 5 to 10% of the agency's budget, and he has bigger plans for the future. "The goals are 20% in basic research sustained," he says.

Devil in the detail

Lofty goals aside, observers of the directorate say Cohen still has plenty of mundane problems to deal with. Proposals still take up to 90 days to wind their way through the directorate's acquisitions branch, says Jill Hruby, director of homeland security programmes at Sandia National Laboratories in Livermore, California. "It's still a long way from being a well-oiled production," she says.

The directorate also draws regular fire from Congress for management shortcomings — most recently earlier this month , when the House Committee on Science and Technology blasted the directorate for neglecting a radiological lab in Manhattan. Congress is impatiently awaiting a five-year plan, and Wu says that he would also like to see a more comprehensive assessment of how it determines the greatest national risks.

The clock is ticking, says Grossman. Cohen is a political appointee, and his two-year term will end in 2008, when a new president is elected. Given the number of problems facing the directorate, she says, expectations should be realistic. "If Cohen just gets started," she says, "it will be a good use of two years." ■

The rules were proposed in December 2006, but it wasn't until 9 April — when the list of chemicals covered by the regulations was released — that safety officers realized that most mid-sized and large universities would have to complete the inventory, called 'Top-Screen'. Many universities estimate that doing so will take thousands of hours, says Peter Reinhardt, co-chair of the government relations committee at the Campus Safety Health and Environmental Management Association in Itasca, Illinois.

"The DHS spent a great deal of time writing these rules with the chemical industry," says Reinhardt. "If it wants to make appropriate rules for colleges and universities, it should spend just as much time with us."

Russ Knocke, a spokesman for the department, is unsympathetic to complaints that the Top-Screen will be onerous. "It comes down to the reality that we do live in a post-9/11 world," he says. "I don't think anyone wants to find themselves in the position where having articulated a desire not to cooperate there is then an incident or an attack."

A public-comment period ends on 8 June, and final rules are likely to be finalized soon after that. Institutions would then have 60 days to complete their Top-Screens.

Emma Marris

Science high on French political agenda

Science has made an unexpectedly strong showing in the government of François Fillon, prime minister of France's newly elected president, Nicolas Sarkozy. The environment and energy, in both the French media and government, have become *à la mode*.

Ecology and sustainable development, long relegated to puny ministries, have been propelled to a top-rank superministry. And at its helm is a political heavyweight — Alain Juppé, a former prime minister and foreign minister. The ministry will have responsibility for the huge sectors of transport, urban and rural planning, energy policy, and other ecological areas such as biodiversity, water and pollution.

P. ANDRIEU/AFP/GETTY

Grand gestures: Alain Juppé has taken the reins at France's new superministry for ecology and sustainable development.

Science and higher education have been granted a full-blown ministry, too, headed by Valérie Pécresse. The move was by no means guaranteed, given that Sarkozy has halved the number of ministers in his government to 15, and that the previous government allotted these sectors only junior status. Bernard Kouchner, co-founder of the medical humanitarian aid group Médecins Sans Frontières and a member of the opposition socialist party, has been appointed foreign minister. And Fillon, a moderate conservative, is himself no stranger to these areas: he served as research and higher-education minister in 1993–95 and 2004–05.

Much of the credit for getting the environment so high on the political agenda must go to Nicolas Hulot, a highly popular TV environmentalist, journalist and writer. During the presidential election campaign, the Nicolas Hulot Foundation asked candidates to sign up to a ten-point ecological pact pledging a radical revision of policies, including energy, transport and agriculture, and to address climate change, species extinction, pollution and other environmental issues. In the face of Hulot's massive popular support (polls showed that had he run, he could have won as much as 10% of the vote), candidates, including Sarkozy, queued up to sign.

It is too soon to say how the creation of France's superministry will translate into actions. But the scale of the government's environmental commitments on paper is "historic", says Yvon Le Maho, a biodiversity researcher at the Hubert Curien Multidisciplinary Institute in Strasbourg. Le Maho, along with several green non-governmental organizations, attended a planning meeting with Sarkozy and Juppé on 21 May to help hammer out a comprehensive five-year plan for the environment. "To be at such a meeting just days after Sarkozy took office was completely surreal," he says.

The government's surprising ecological bent could also make for an interesting G8 meeting, which is scheduled for 6–8 June in the more traditionally green Germany. ■

Declan Butler

Complex set of RNAs found in simple green algae

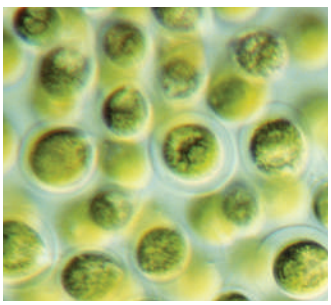
A class of RNA molecule, called a microRNA, has been found in a unicellular green alga. The discovery, made independently by two labs, dismantles the popular theory that the regulatory role of microRNAs in gene expression is tied to the evolution of multicellularity.

The researchers confirmed the existence of dozens of microRNAs in the genome of the green alga *Chlamydomonas reinhardtii*, and think that hundreds are likely. The finding is as startling as the discovery ten years ago that the nematode *Caenorhabditis elegans* has 19,000 genes, just 1,000 short of the human count, says Gregory Hannon from Cold Spring Harbor Laboratory in New York and a co-author of one of the studies (T. Zhao *et al. Genes Dev.* 21, 1190–1203; 2007). "People were shocked that the complexity of the genomes in these simpler creatures was similar to our own," he says. Now it seems that the RNA in simple unicellular organisms could be as complex as that in higher creatures.

The second study, which independently came to the same conclusion, appears online in *Nature* this week (A. Molnár *et al.* doi:10.1038/nature05903; 2007).

Ever since the discovery of RNA interference — the selective blocking of gene expression by small RNAs — biologists have identified a growing family of these tiny molecules in eukaryotic organisms. But hunts in yeast and in the protist *Tetrahymena* have yielded fewer types of RNA molecule than found in plants and animals. MicroRNAs were never found.

This, combined with the fact that RNA sequences differ between plants and animals, helped give rise to the idea that microRNAs evolved independently in plant and animal lineages as parts of complex regulatory mechanisms associated with multicellularity. Now it seems that these molecules may predate that evolutionary development.



M.I. WALKER/SPL

MicroRNAs found in green algae hint at the organism's complexity.

"It shows how basing conclusions on studies of just one or two model organisms can really lead you astray in terms of how you think about evolutionary processes," says Jim Umen from the Salk Institute in La Jolla, California.

A combination of factors led to the recent findings: high-throughput sequencing can now sieve through thousands of RNA molecules in search of microRNA. And having a nearly completed map of the genome for the alga means that its sequences can be interpreted relatively easily.

Nobody knows why such a simple organism needs microRNAs, nor how or when they first appeared. But researchers say they may help *C. reinhardtii* adapt to extremely diverse environments.

Whatever their role, their presence indicates that microRNAs could be much more ancient than previously thought; they might have persisted for more than a billion years. ■

Lucy Odling-Smee

Abrupt end for Australian foray into Singapore

An Australian university's bold experiment in exporting higher education to Singapore has ended much more quickly than anyone could have imagined.

On 28 June, the University of New South Wales Asia will be calling it quits after just one term. A university statement last week blamed low student numbers — fewer than half the expected 300 students had been enrolled.

The experiment aimed to provide research-based degrees, unlike some other 'imported' universities that offer just an English-language education. The University of New South Wales spent Aus$17.5 million (US$14.4 million) setting up the venture. It reportedly attempted to negotiate downsizing with the Singaporean government, but the proposal was rejected.

Singapore has been trying to attract international educational and research institutions to fuel its dream of becoming a cutting-edge research power. But last year, Johns Hopkins Medicine in Singapore announced the closure of its research facility among mutual allegations of broken promises (see *Nature* 442, 493–493; 2006).

Britain launches public debate on nuclear power

The UK government has stopped short of giving its full backing to new nuclear power stations in its latest proposals for the country's energy strategy. In a policy outline published on 23 May, it opened a public consultation on the issue — including several potential sites for new nuclear facilities, which the government is thought to view as essential for ensuring energy security while meeting its self-imposed target of cutting carbon emissions by 60% by 2050.

The document also sets out details of a mandatory carbon-trading scheme for large businesses such as banks and supermarkets. And it describes plans to build the world's most advanced carbon-storing fossil-fuel plant, to be opened by 2014, which will in theory capture 90% of its own greenhouse emissions.

L. PITARAKIS/AP

The UK government sees nuclear power as key.

But the news came as oil giant BP announced it was pulling out of plans to build a major carbon-capture plant in Peterhead, Scotland. The company said it did not want to delay the plant further by waiting for the bidding competition for funding to start as laid out in a timetable in the new white paper.

Fishers feel the long arm of the law

Crime-fighting, fingerprint-tracking tools have been turned on a new target in New York state: the small, weasel-like mammals known as fishers (*Martes pennanti*).

Researchers say they can use the tools to compare footprints from various fishers (pictured) as a non-invasive method of counting how many of the creatures live in a particular area. The technique could work better than monitoring skin or fur markings to track other rare species in the wild, and it is cheaper than DNA fingerprinting, the group says (C. J. Herzog *et al. J. Wildl. Mgmt* **71,** 955–963; 2007).

But the animals may yet confound the researchers. At six baited stations spaced throughout the Adirondack Mountains, most of the footprints came from the sooty paws of a single fisher.

WILDLIFE CONSERVATION SOC.

NIH calls halt to breeding chimps for research

The US National Institutes of Health has indefinitely extended a moratorium on breeding captive chimpanzees for research.

The moratorium has been in place since 1995, and affects only the 650 chimps owned by the agency's National Center for Research Resources. John Harding, director of primate resources for the centre, says that it simply costs too much to care for chimps — $300,000 to $500,000 over a single animal's lifespan.

Animal-rights groups welcomed the decision. But many researchers are disappointed. Some have argued that the glut of genomic data arising on chimps and other non-human primates will be useless without further studies on the biology of the animals. They will now have to hurry to find out what they can: in March, a group of external advisers said that the US population of research chimpanzees will die off within 30 years.

Radiation-ecology lab prepares to close its doors

A US lab that specializes in the ecology of radioactive waste looks set to close in the near future. "We are probably the last laboratory with expertise in radiation ecology at a time of nuclear renaissance," says Paul Bertsch, director of the Savannah River Ecology Laboratory near Aiken, South Carolina. The lab sits on the Department of Energy's Savannah River Site, which is home to a large stockpile of nuclear weapons and waste. For 56 years the department has funded research at the lab into how various toxic substances, including radioactive materials, interact with the environment.

But in recent years the agency has encouraged the lab to be less reliant on energy-department funds. And earlier this month the department announced that only 6 of about 20 proposed research activities were worthy of funding.

The House Committee on Science and Technology is investigating the cutback in funds, which committee staffer Dan Pearson calls "incomprehensible".

Gotham prize hedges bets for fight against cancer

US hedge-fund managers have teamed up with scientists to launch a competition for the next big idea in cancer research.

Applicants must first be accepted by an Internet-based club (www.gothamprize.org), whose membership will be vetted by a scientific advisory panel that includes cancer experts such as Bert Vogelstein of the Johns Hopkins Medical Institutions in Baltimore, Maryland. Members can then submit their idea — in fewer than 1,000 words — for a research project in basic cancer research, or in cancer diagnosis, prevention or treatment.

The person whose idea is judged to have the greatest potential will win US$1 million, even if they will not themselves be carrying out the research to test it. Another prize of $250,000 will be given in paediatric oncology. The ideas that emerge will be shared with other cancer-research funders.

The organizers say that current funding opportunities tend not to support untested ideas and that the annual Gotham prize will help fill this gap.

BUSINESS

Algae bloom again

A handful of pioneers are trying to bring algae-based biofuels back from a near-death experience. **Amanda Leigh Haag** reports.

On a spring afternoon in Fort Collins, Colorado, a bath of lime-green algae is glinting in the sunshine. Filling an 18-metre-long tank that sits in the parking lot of a disused power station, the bubbling liquid consists of millions of microscopic single-celled photosynthetic organisms. But for local company Solix Biofuels, this is more than a warm bowl of slime: it's a photobioreactor, which could soon serve as a viable energy source — and as a carbon sink, to boot.

Surging oil prices and a shortage of biofuel feedstocks are reviving interest in making fuel from algae. A barrage of tax credits and subsidies has already spawned a boom in facilities to make biofuels from corn (maize), rapeseed (canola), soya and oil palms — and hiked up the prices of these commodities. But advocates of algae say they have attributes that should make them strong contenders in the long run.

According to figures compiled by the Global Petroleum Club, a business community for leaders in the oil, gas and energy industries, soya typically produces 450 litres of biodiesel per hectare per year; canola 1,200 litres; and oil palm 6,000 litres. Researchers predict that a hectare of algae could produce 90,000 litres of biodiesel, and have the potential to go even higher. They can also be grown on land that is agriculturally barren. "Algae offer orders of magnitude greater resource potential for natural oils than any terrestrial crop," says John Sheehan, an energy analyst with the National Renewable Energy Laboratory (NREL) in Golden, Colorado.

SUNSOURCE/BIODIESEL PROJECT

"What happens now if people are all geared up for biofuels, but the market's not there?"

— Doug Henston

In addition, algae consume carbon dioxide as they grow, so they could be used to capture CO_2 from power stations and other industrial plant that would otherwise go into the atmosphere.

In the short term, Solix and the handful of other companies working on algae want to produce oils for biodiesel, replacing a significant proportion of the diesel fuel that currently serves about one-third of transport needs in the United States. Eventually, they plan to produce ethanol from the starchy residues as well.

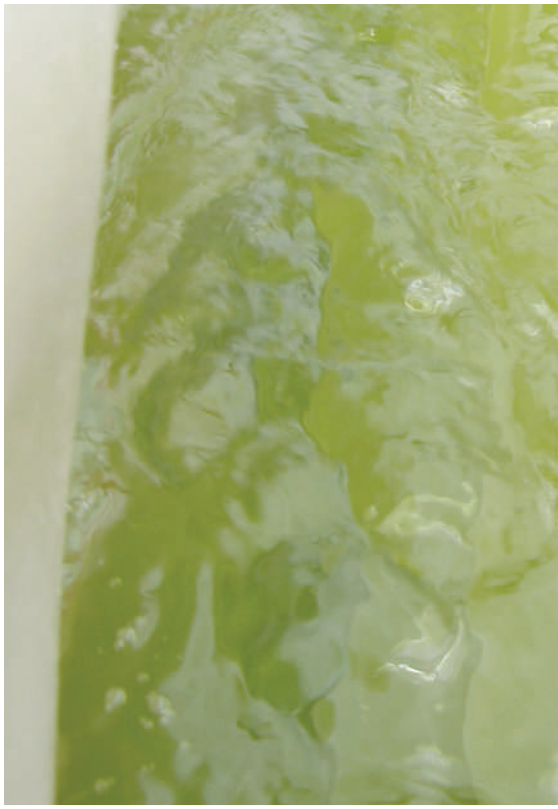
Progress on getting energy from algae pretty well ground to a halt in the United States in 1996 when the NREL abandoned an 18-year-old programme on it. NREL researchers sifting through thousands of algal species had identified 300 that promised the highest oil yields. But they also pinpointed major roadblocks to the technology — such as the difficulty in reliably replicating laboratory growth conditions for algae on a larger scale in the field.

Algae garner energy from sunlight, water and CO_2 through photosynthesis. They reproduce quickly and can be harvested day after day. Each species produces different ratios of oils, carbohydrates and proteins. Oils squeezed from algae can be turned into biodiesel, and the rest of the biomass can be converted into ethanol and animal feed, say advocates. Whereas oil palms — currently the leading producer of oil for biodiesel — typically yield 20% of their weight in oil, some algae yield more than 50%.

Doing all this on a commercial scale isn't straightforward, however. Shaine Tyson, chief executive of Rocky Mountain Biodiesel Consulting in Glenwood Springs, Colorado, is a veteran of NREL's algae programme. "When we actually tried to increase the oil content, we would reduce the total mass yield," Tyson recalls. Some published projections for algae's potential are based on the false assumption that algae will have equally high production rates all year round, she adds.

Algae can also be picky: too much direct sunlight can kill them; temperature must be held steady; overcrowding will inhibit their growth; and the 'waste' oxygen they produce must be continually removed from the water. Even the bubbles of CO_2 can rupture algal cells. Open algal ponds are subject to evaporation and rainfall, which cause salinity and pH imbalances. And local species of algae often overgrow the desired strain — a problem that the NREL never solved.

To address some of these challenges, Solix grows algae in enclosed chambers made from plastic sheeting. But some companies, such as LiveFuels in Menlo Park, California, see open-pond systems as the only hope for keeping capital costs low enough to be commercially viable.



S. JAFFE

Solix has partnered with the New Belgium Brewery and Colorado State University, both in Fort Collins, and plans to start operating a small-scale commercial plant this autumn that will siphon off the CO_2 generated by fermentation in the brewery. Eventually, Solix would like to link up with the power industry: it calculates that a 1-gigawatt coal-fired power plant, which produces some 6 million tonnes of CO_2 annually, would need 16,500 hectares of algal greenhouses to soak up 90% of its CO_2 emissions.

But for now, the algae companies see biodiesel as their main objective, with carbon capture as a bonus. "It's a huge benefit, but it happens not to be much of an economic driver," says Bryan Willson, chief technology officer at Solix. "Currently, the economics of fuel production far outweigh those of carbon capture," he says. "That could change dramatically if there are structural changes made to the carbon regulations in the United States."

LiveFuels has raised $10 million so far and would like to raise another $35 million to set it on its way to its target of making biocrude — the precursor of biodiesel — from algae for $1 per gallon by 2010. "We don't know whether

Green gold: algae could produce biodiesel much more efficiently than conventional crops.

that's achievable, but it is our goal," says Lissa Morgenthaler-Jones, its chief executive. GreenFuel of Cambridge, Massachusetts, has so far raised $20 million from venture-capital groups that support early-stage companies and 'disruptive' technologies.

Solix has solid backing from a local private investor, and says it plans to develop its technology as far as it can on its own before seeking venture capital. Solix believes it can build a system that's competitive on a small commercial scale with between $5 million and $15 million, and says it has sufficient backing to do this.

Solix chief executive Doug Henston warns, however, that it's important not to overstate the case for the technology. He prefers to take a low-key approach: "Biofuels are in a great position to fall on their face," he says, noting President George Bush's pledge to replace 35 billion gallons of petroleum with ethanol and biodiesel by 2017. "What happens now if people are all geared up for biofuels, but the market's not there?"

And all agree that the key to algae's success will be getting costs down. "The challenge is coming up with economical systems," says Al Darzins of the bioenergy centre at the NREL, which is now collaborating with LiveFuels. "When it comes down to it, you're producing a commodity that has to be dirt cheap." ■

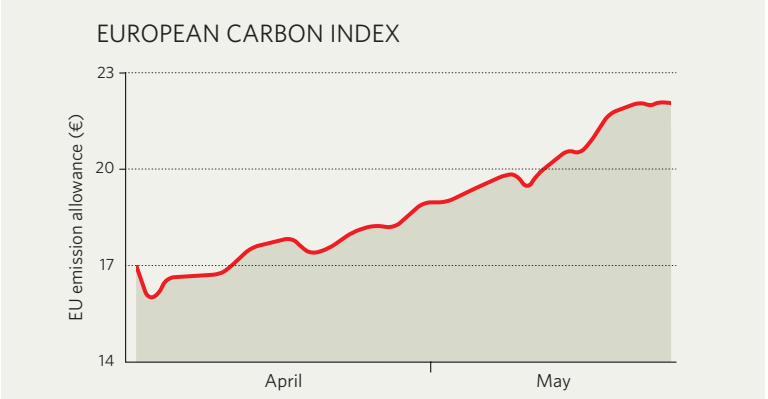
IN BRIEF

ENVIRONMENTAL UPTURN US engineering firm General Electric (GE) says it has doubled sales of environmentally friendly products, to $12 billion annually, in the two years since it announced that it was 'going green' (see *Nature* 435, 410–411; 2005). At an event in Los Angeles with California governer Arnold Schwarzenegger, GE chief executive Jeffrey Immelt said that the company was on track to meet its target of $20 billion in sales of such products by 2010. The goods in question range from energy-efficient light bulbs to railway locomotives — GE also introduced what it says is the world's first diesel-electric hybrid locomotive at the event.

COPYCAT CATCH Sun Pharmaceuticals, India's biggest drug maker, has confirmed that it will purchase Israeli generics maker Taro for US$454 million. The purchase, which reflects an increasingly acquisitive streak among Indian drug makers, will help the Mumbai-based company make inroads in the United States, where Taro has most of its sales. The Israeli company has been in financial difficulties, but Sun's shares rose by about 5% when the deal was announced on 21 May.

GROUNDED NASA Computer company Hewlett-Packard has won a mammoth contract, worth up to $5.6 billion, from NASA, to supply computers and other equipment to parts of the federal government. The unusual contract enables any government agency to buy the kit within a framework agreed between the space agency and the computer maker, which is based in Palo Alto, California. It's a follow-on from two far smaller contracts that have run since 1992.

MARKET WATCH



SOURCE: EEX

As the European Commission implements more stringent national caps on carbon dioxide emissions this spring, demand is rising on the emissions markets for the allowances that let companies emit more CO_2.

Several million allowances are traded daily at five European carbon exchanges, including the European Energy Exchange (EEX) in Leipzig, Germany (above).

Many energy-supply companies in the European Union (EU) are snapping up allowances now, in the expectation that they will cost more by next year. The price of an allowance to emit one extra tonne of CO_2 in the second phase of the European emissions-trading system, from 2008 to 2012, has almost doubled since February, reaching a year high of €22 (US$30) on 22 May.

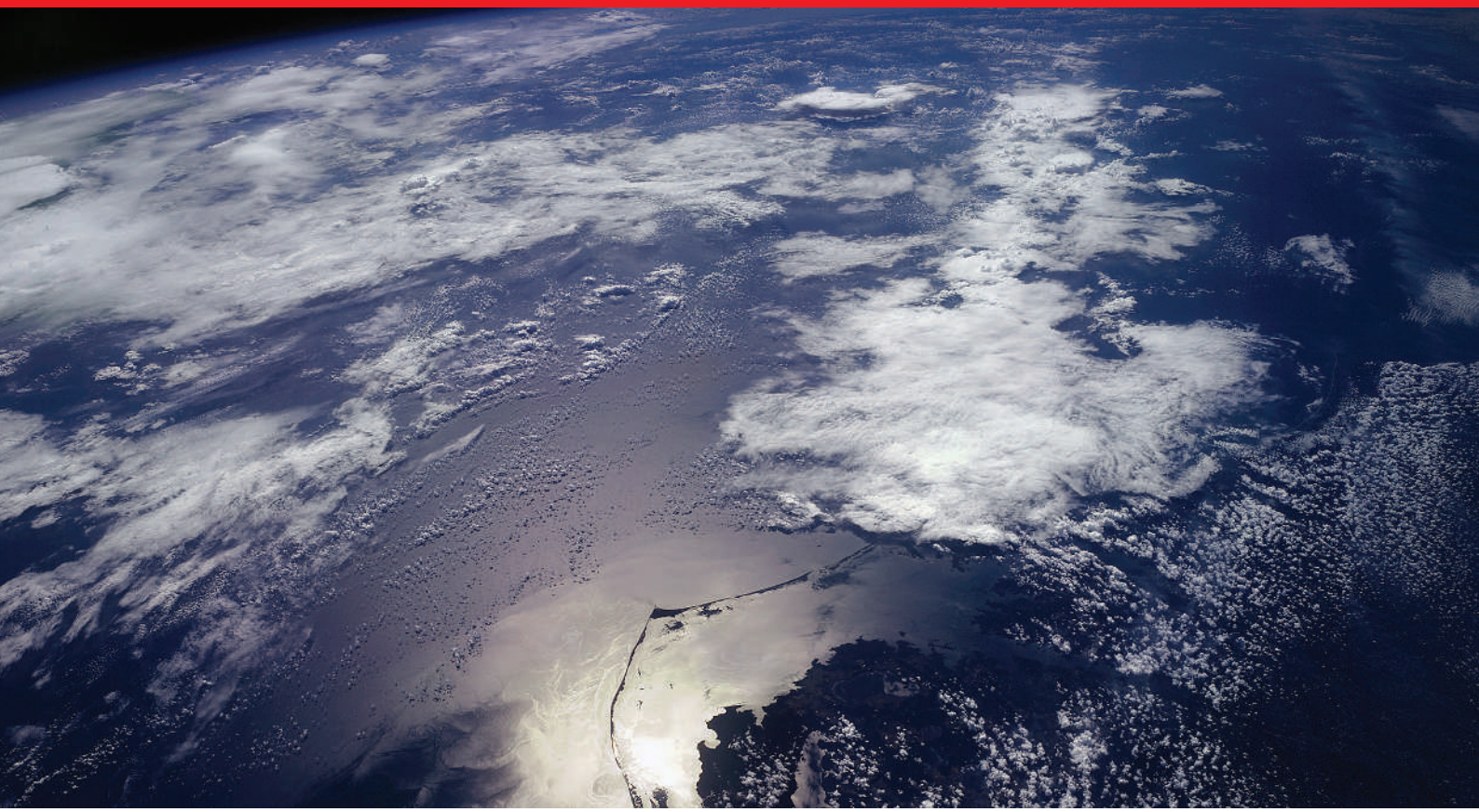
EU emissions trading, introduced in 2005, seems to have got over an early phase of pronounced, and at times worrying, turbulence, analysts say (see *Nature* **441,** 405; 2006). Thanks to the reduced caps for the second trading period, the market is now unlikely to collapse, according to one senior market-watcher based in Amsterdam.

"There is a lot more clarity now," agrees Milo Sjardin, who watches carbon markets for New Carbon Finance in London. "Uncertainty hasn't gone away entirely, but the market is definitely more robust than it was."

There is not yet enough liquidity on the market to prevent single, large transactions from moving the price up or down. But brokers are confident that the price will continue to rise moderately through the summer, perhaps levelling off towards the end of the year when power companies have signed their contracts for selling electricity in 2008. ■

Quirin Schiermeier

NATURE|Vol 447|31 May 2007



CORBIS

Churn, churn, churn

How the oceans mix their waters is key to understanding future climate change. Yet scientists have a long way to go to unravel the mysteries of the deep. **Quirin Schiermeier** reports.

The absence of data is not always a bad thing; it can sometimes make a job a lot easier. In decades past, for instance, climate modellers simply didn't worry too much about how turbulent processes in the ocean (known as ocean mixing) affected the outputs of their models. Not knowing how ocean mixing worked may not have been intellectually fulfilling, but it certainly made the models simpler to run.

But data are now beginning to emerge — and they show, peskily, that mixing is not so simple after all. Information flooding back from instruments such as microstructure profilers, which are towed behind ships to gather data on factors such as temperature and conductivity, has started to quantify exactly how parcels of water mix with each other. In particular, the amount of mixing seems to be much less than models of the ocean would suggest, especially given the temperatures measured at the ocean's depths. Some of the mixing is missing.

As a result, ocean mixing is an increasingly hot topic at conferences. "The field is changing quicker than many would have thought," says Raffaele Ferrari, an oceanographer at the Massachusetts Institute of Technology (MIT) in Cambridge. "It seems unlikely that we know the full story."

The importance of mixing is that it helps the oceans move heat from A — normally near the Equator — to B — which will typically be nearer one of the poles. The best known mechanism is the global conveyor belt, or thermohaline circulation; warm waters move north through the Atlantic before becoming cool and salty enough to sink to the ocean bottom and flow back south, from where they are distributed worldwide[1]. The energy involved is prodigious: 2,000 trillion watts, or two petawatts, some 200 times the rate at which mankind uses energy, and a significant chunk of the energy flow needed to drive Earth's climate.

Since early in the twentieth century it has been known that, without mixing, such a system would stall. The depths would fill up with cold water that, because it was dense, would not resurface, leaving a thin warm layer on top. Warmth is needed to add buoyancy to the bottom waters, and that means mixing them with the waters above. But no one can say for sure where the energy needed that drives this mixing comes from. Some have suggested plankton as paddles (see 'Could tiny creatures stir the whole ocean?'). Others have called down the Moon in support of their theories, proposing tides as the driving force.

> **"It seems unlikely that we know the full story."**
> — Raffaele Ferrari

The prevailing trade winds between the tropics of Cancer and Capricorn were perhaps the most obvious candidates for stirring the oceans, in principle providing sufficient energy to create a well-mixed upper layer in the tropical oceans. But the amount of mixing near the ocean's surface, as observed by techniques such as microstructure profilers and tracer experiments, is not as high as theoretical calculations would suggest, and not enough to keep the global circulation going.

Stormy weather

One recent idea is to look not at the regular winds that are the sailor's friend, but the rare ones that spell his doom — specifically, hurricanes. On page 577 of this issue, modeller Matthew Huber of Purdue University in West Lafayette, Indiana, presents evidence supporting the notion that tropical cyclones play an important part in ocean mixing. Huber's work follows on from a 2001 study that suggested much the same thing[2]. In that earlier work, Kerry Emanuel, a hurricane researcher at MIT, proposed that tropical cyclones mix the upper ocean so well that they are responsible for a staggering amount of heat transfer: 1.4 petawatts. As most other researchers had ignored the role of cyclones

completely, perhaps thinking them too sporadic to have such a sustained effect, the idea was met with some scepticism. The new study, based on more and better observations, dials this back to an estimated 0.26 petawatts per year — a number that Emanuel says is still "in the right ballpark" for the level of mixing he would expect from cyclones.

Even with the lower mixing rates cited in Huber's study, the very idea that cyclones could substantially contribute to ocean mixing is so surprising that many are still likely to reject it. "At least in the present climate the effect is unlikely to be very large," says Ferrari. But that would not necessarily be the case in the climates of tomorrow (or, for that matter, the climates of yesteryear — see 'Lessons from the past').

What is thought provoking about Emanuel's ideas is that he has also shown that high sea-surface temperatures could push up the strength of tropical storms. Others have speculated that global warming might also make storms more frequent. If cyclones are indeed playing a key role in ocean mixing, then increases in their power and frequency could change that role, and thus the climate. Hurricanes could matter far farther afield than the coasts they batter — their effects could be worldwide. There have long been debates on what global warming means for hurricanes; now we may see one on what hurricanes mean for global warming.

T. HOEPKER/MAGNUM

Mix it up: cyclones might help to transfer heat from the Equator to the poles.

Tales from topographic oceans

Another source of mixing that was long ignored but is now increasingly accepted is down to the Moon and the tides, and the effect they have, not at shores, but in the oceans' depths. In the 1960s and 1970s, conventional oceanographic wisdom held that the considerable energy represented by the ebbing and flowing of tides was dissipated entirely through friction in the shallow coastal areas above continental shelves. But then pioneering oceanographers Walter Munk of the Scripps Institution of Oceanography in La Jolla, California, and Carl Wunsch of MIT proposed that tides penetrate all the way into deep waters, where their flow across the rough seafloor topography helps the water to mix and overcome the forces of gravity[3]. Studies from the Topex/Poseidon satellite that maps sea-surface elevation have since confirmed that tidal energy does indeed get dissipated in the deep ocean[4].

Other experiments have lent further support to this idea. The Hawaii Ocean Mixing Experiment has quantified the amount of mixing over undersea ridges[5]; it found tides with amplitudes of up to 300 metres sweeping across the flanks of one ridge. Separately, the AnSlope project off Antarctica has looked at the continental shelf surrounding Antarctica and how it affects ocean mixing throughout the Southern Ocean. These projects have shown that most of the mixing takes place at the ocean's boundaries, from the drop-off of the continental slope

Could tiny creatures stir the whole ocean?

Billions of animals such as krill (pictured) and plankton constantly plough through the ocean's waters, from the tropics to the poles. Could their swimming actually help mix the oceans? "The idea is so fascinating, some would rather hope it is true," says Carl Wunsch, an oceanographer at the Massachusetts Institute of Technology in Cambridge.

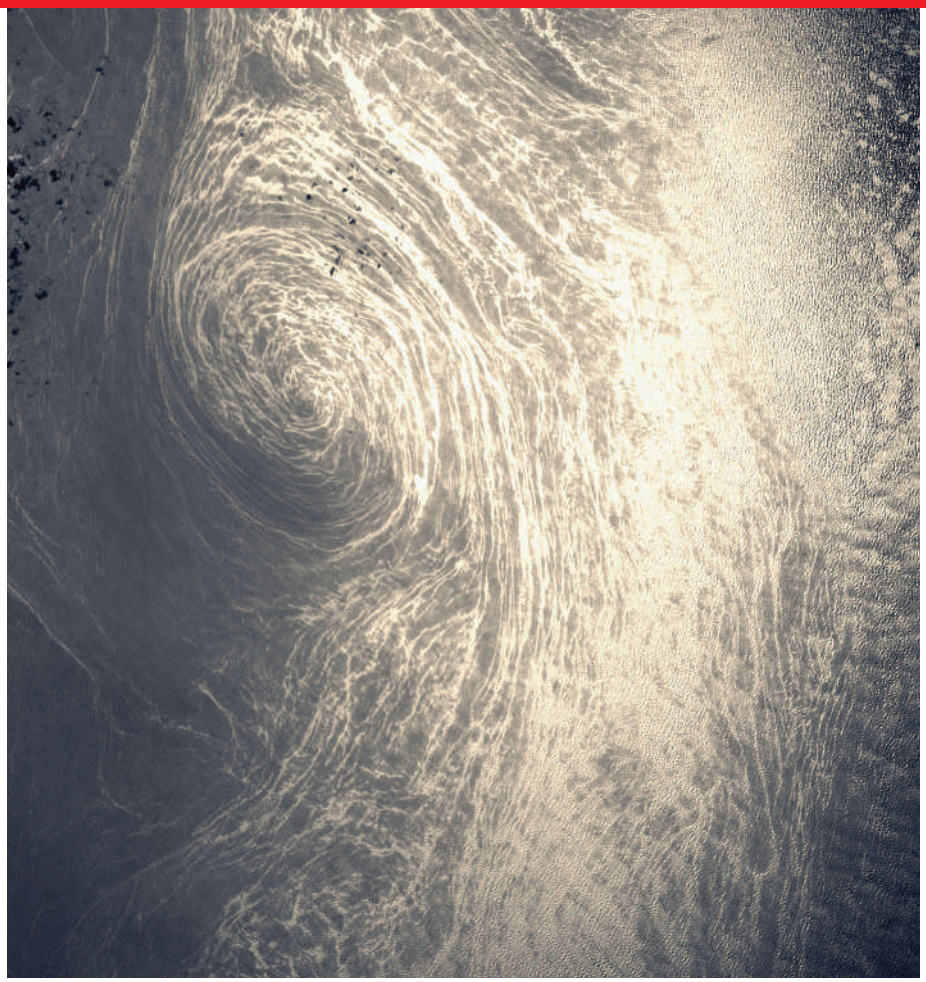
The idea of 'biomixing' recently underwent something of a revival. Scientists reported last year[8] that large schools of krill stir a well-stratified fjord along the coast of British Columbia, Canada. Another study[9], also published last year, calculated that the overall amount of energy supplied by the swimming motion was substantial. Marine life, the authors suggest, could supply up to half of the mechanical energy required worldwide to mix cold waters to the surface.

If it is real, biomixing will be a serious complication for ocean modellers. They will "need to start thinking about the fluid dynamics of biology," says Wunsch. "That's a tough one."

K. AITKEN/V&W/IMAGEQUESTMARINE.COM

However, the jury remains out as to whether biomixing really has that great an effect. Even if marine animals do dissipate large amounts of energy through their swimming, the resulting mixing of the waters may be small. The problem is that most or all of the energy from moving critters is converted into frictional heat, rather than being available for shifting heavy water.

"Mixing efficiency has simply been ignored," says Andy Visser, a marine physicist at the Danish Institute for Fisheries Research in Charlottenlund. "Friction releases heat — that's why basic physics should tell you that there can never be much biomixing." **Q.S.**



CORBIS

Eddies such as those seen in the Gulf Stream may play a part in mixing the oceans' water.

to the tall mid-ocean ridges that run along the middle of the sea floor. Mixing doesn't occur everywhere to the same extent. It happens more around and above rough mountainous regions on the sea floor than, say, in the vast abyssal plains. This variability from place to place further complicates modellers' efforts to capture the mixing accurately.

As always, new observations could help. Researchers are unleashing more microstructure profilers, along with remote sensing of sea-surface temperatures and sea-level anomalies, to keep trying to track down the missing mixing. Among other things, they need better to understand the role of 'mesoscale' eddies, some 50 to 100 kilometres across, that swirl around some parts of the oceans mixing water and affecting biological productivity[6].

Meanwhile, modellers are far from keeping up. Until now they have mostly pretended that mixing occurs uniformly throughout the ocean, and everywhere at the same rate. No one believes that is how the real ocean behaves; few, if any, think it is even a reasonable approximation for a functional climate model. "It's a really serious weak point," says Olivier Marchal, an ocean modeller at the Woods Hole Oceanographic Institution in Massachusetts.

Without capturing the complexities of mixing better, the models will be hard put to capture how climate can be expected to change thanks to the greenhouse effect. For example, most researchers expect the oceans to become less dense.

But that could change, depending on a key mixing parameter, called diffusivity, that most computer models have so far regarded as fixed. A changing mixing in the models could ultimately alter our view on key components of the climate system, such as the thermohaline circulation; in some models increased mixing could render it more powerful[7]. Researchers would be better off accepting this idea, many say, and moving forward with new ways to quantify the problem. But working the fine details of mixing into global models will remain a tall order.

With so many new findings, those who study ocean mixing are scrambling to incorporate them all. And scientists may soon have a rule book by which to conduct their future research. Robin Muench, an oceanographer with the Earth and Space Research Institute in Seattle, Washington, leads a working group set up by the International Council for Science. Its goal is to identify the rate of mixing in the oceans as a whole as well as in specific regions where mixing is known to occur (or not).

In July, the group is slated to deliver its final report in Perugia, Italy, which should help clarify the research agenda. And with that, the modellers will have their next couple of years' work laid out for them. ■

Quirin Schiermeier is a correspondent for *Nature* in Munich.

Lessons from the past

J. SCHMALTZ/MODIS/NASA

Hurricanes churning up the oceans could provide an elegant, albeit speculative, explanation for one of the most striking mysteries in Earth's climate history.

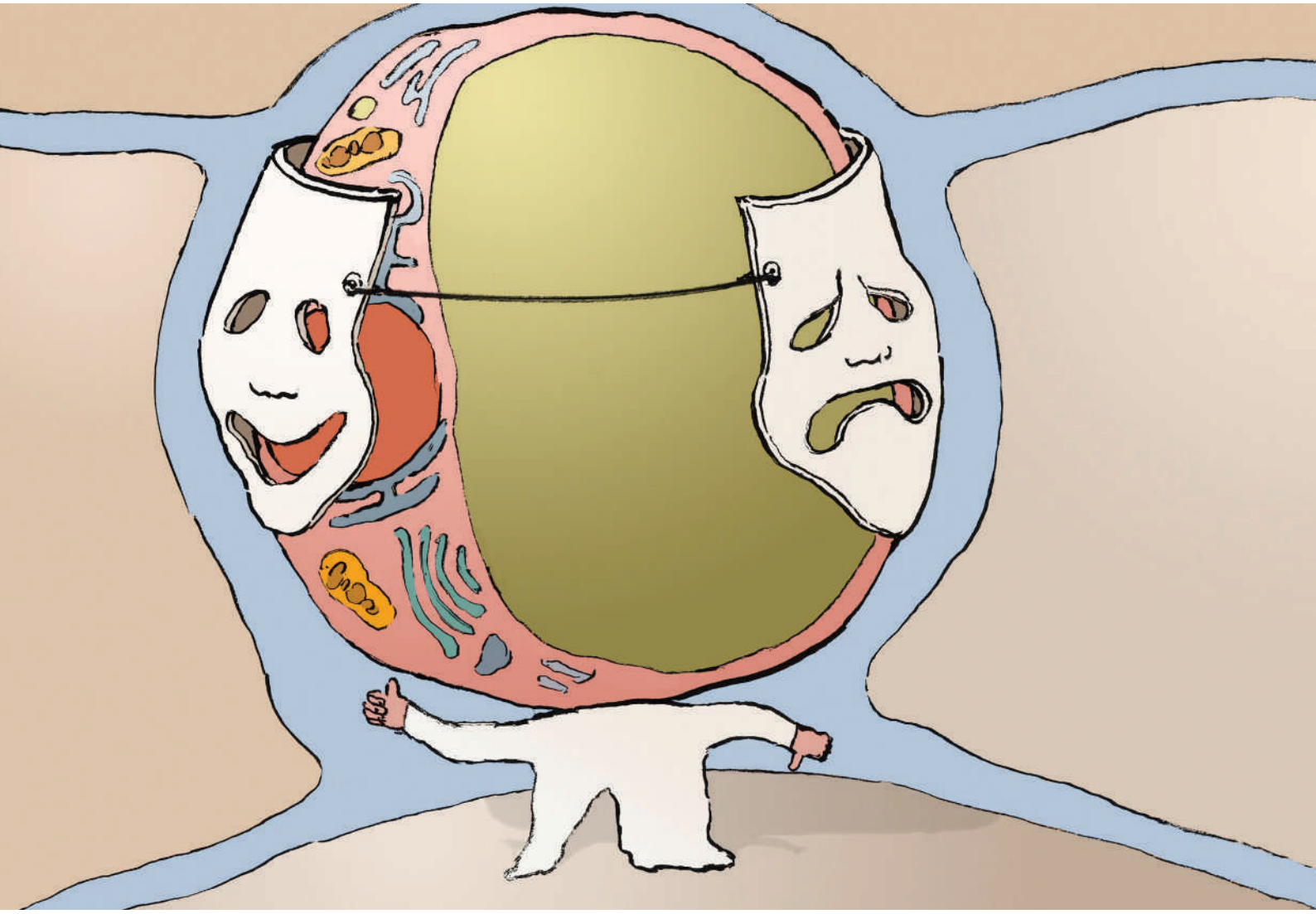
During the Eocene, some 50 million years ago, the planet was much warmer than it is today. Sediments pulled from the bottom of the Arctic Ocean suggest that the North Pole enjoyed an almost Mediterranean climate, with summer temperatures as warm as 24 °C (ref. 10).

Climate models provide no adequate explanation. And the extraordinary temperatures in the Arctic imply that the tropics must have been excessively hot.

But could intense cyclones have helped to cool things down? Many scientists, including Kerry Emanuel of the Massachusetts Institute of Technology in Cambridge, have speculated about the possibility. They think that strong and frequent cyclones may have enhanced mixing in the upper layers of the ocean, helping to transport heat from the tropics to the poles.

But so far there is little evidence that warmer climates produce more tropical storms. The search for storm signatures in geological deposits has produced ambivalent results at best, says Emanuel. **Q.S.**

1. Schiermeier, Q. *Nature* **439,** 256–260 (2006).
2. Emanuel, K. *J. Geophys. Res.* **106,** 14771–14782 (2001).
3. Munk, W. & Wunsch, C. *Deep Sea Res.* **45,** 1977–2010 (1998).
4. Egbert, G. D. & Ray, R. D. *Nature* **405,** 775–778 (2000).
5. Rudnick, D. L. *et al. Science* **301,** 355–357 (2003).
6. McGillicuddy, D. J. *et al. Science* **316,** 1021–1026 (2007).
7. Nilsson, J., Brostöm, G. & Walin, G. *J. Phys. Oceanogr.* **33,** 2781–2795 (2003).
8. Kunze, E., Dower, J. F., Beveridge, I., Dewey, R. & Bartlett, K. P. *Science* **313,** 1768–1770 (2006).
9. Dewar, W. K. *et al. J. Marine Res.* **64,** 541–561 (2006).
10. Brinkhuis, H. *et al. Nature* **441,** 606–609 (2006).



J. KAPUSTA

THE TWO FACES OF FAT

No longer viewed as inert packets of energy, fat cells are two-faced masterminds of metabolism. **Kendall Powell** weighs up the differences between 'fat' fat cells and thin ones.

Unwanted, unloved, yet often overabundant, few have much regard for fat. Scientists, too, long thought of fat cells as good-for-nothing layabouts unworthy of attention; containers stuffed with energy to be released at the body's command. So stuffed, in fact, that many other parts of the cell were thought too squeezed to function.

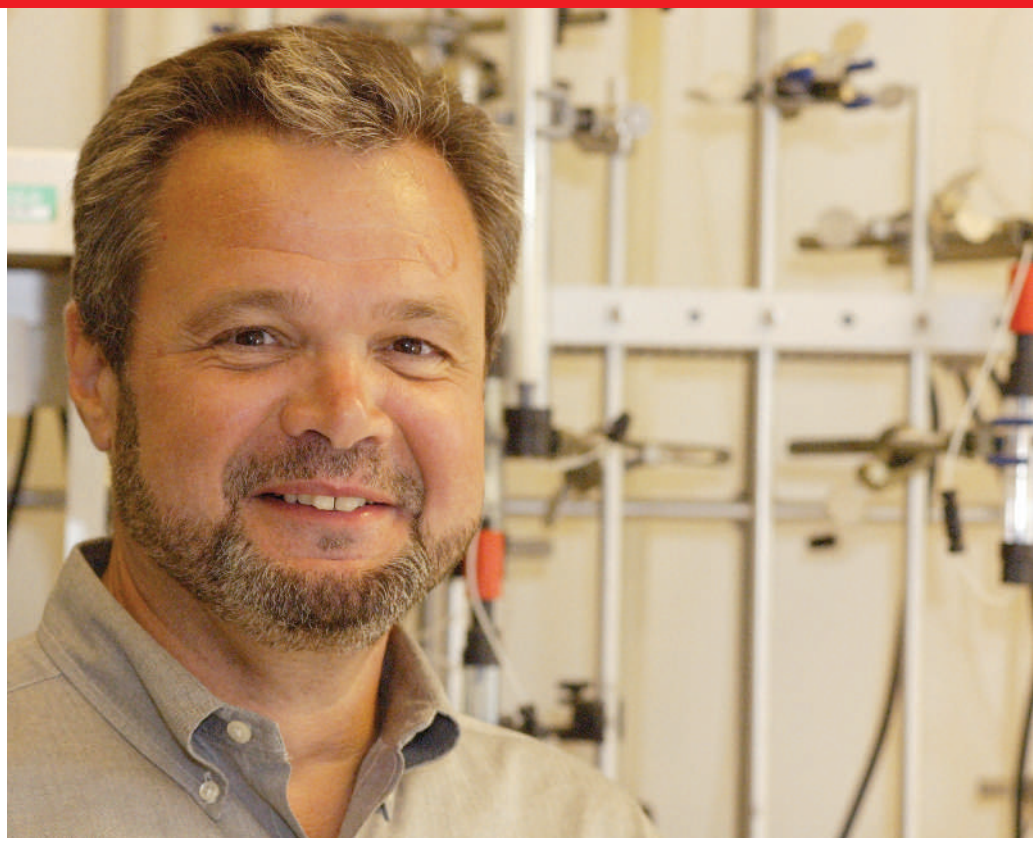
So when, in the early 1990s, graduate student Gökhan Hotamisligil at Harvard Medical School in Boston caught fat tissue doing something biologically remarkable, at first he did not believe his own data. He repeated the work many times, but it always came out the same: fat from obese mice was producing TNFα — the hot inflammatory molecule of the day because of its role in autoimmune disorders such as arthritis. After he and his colleagues published their observation in *Science*[1], others in the field remained sceptical. Hotamisligil says he was invited to speak at meetings "for entertainment purposes".

Since then, fat cells have had an image change. This started with the 1995 discovery that fat secretes leptin, a hormone that tells the brain "I'm full, stop eating". In retrospect, it makes sense that fat should tell the body how much energy it is storing and how much more to take in. But when it came to obesity-related problems such as type 2 diabetes and cardiovascular disease, fat was still not seen as an active player. These conditions were thought to be caused by an excess of nutrients from overeating, or a glut of fatty molecules spilling out of storage into the bloodstream.

More than a decade on, fat has a higher status. Scientists know that fat cells pump out ten or more molecules called adipokines that carry messages to the rest of the body. And 'fat' fat cells — those common in the obese and which are themselves bloated with lipids — send different molecular messages from 'thin' fat cells.

The signals from 'fat' fat are thought to directly promote insulin resistance and to trigger inflammation, which may, in turn, cause type 2 diabetes, cardiovascular disease, increased cancer risk and other obesity-associated problems. This means that it might be possible to treat these conditions without shedding the fat itself. Some remedies might be as simple as using anti-inflammatory drugs that have been around for more than a century; others might involve persuading obese fat cells to behave like skinny ones.

Society may still view fat with resignation or even revulsion, but biologists have moved on. "No one appreciated the higher functions of fat," says Barbara Kahn, a diabetes and obesity researcher at Harvard's Beth Israel Deaconess Medical Center in Boston. "The fat


M. J. MALONEY

Steven Shoelson is trialling a simple painkiller against the inflammation that accompanies diabetes.

cell's got to be put on the map now."

An animal fat cell, or adipocyte, is a giant droplet of triglyceride molecules — each composed of fatty acids and glycerol — plus a life-support system of other organelles squashed to the side. After a meal, fatty acids and glucose enter the blood. Fat cells absorb the fatty acids, and the liver converts excess glucose into more fatty acids, which fat cells take up and store. In obesity, when the body is swimming in excess fats and glucose, fat cells pack more in and expand.

Sugar daddy

Fat and glucose control are linked: overweight people tend to develop insulin resistance and then type 2 diabetes, conditions in which the hormone insulin no longer promotes normal glucose uptake and fails to stem glucose production by the liver. But researchers generally assumed fat was a bystander in these problems. They thought the defect must lie in muscle and the liver, which take up and metabolize the vast majority of blood glucose.

Then, in 2001, Kahn's group showed that fat tissue was managing much of the body's response to insulin. The researchers used genetic engineering to eliminate a glucose transporter molecule from the surface of mouse adipocytes, and found that the animals' muscle and liver cells also became insulin resistant — just as much as those from obese mice[2]. This suggested that obese fat cells make a circulating factor that makes other tissues resistant to insulin.

In 2005, Kahn identified that molecule as RBP4 (ref. 3), which works partly by blocking the action of insulin in muscle and liver. The team went on to show that obese and type 2 diabetic patients have higher levels of RBP4 in their blood compared with healthy controls[4]. "Fat cells are really conducting the orchestra, telling the sugar where to go," says Evan Rosen, who also studies obesity at Beth Israel Deaconess but was not involved in the studies.

Because of its link to human disease, RBP4 is one of the stars of a growing list of molecules secreted at higher levels by 'fat' fat cells compared to 'thin' ones — or vice versa. They include adiponectin, which is the opposite of RBP4 in that it improves the action of insulin, and is secreted in large quantities by normal fat cells but less so by 'fat' fat cells. Last year, a group led by diabetes researcher Philipp Scherer of the University of Texas Southwestern Medical Center in Dallas showed that mice lacking adiponectin suffer severe insulin resistance in the liver[5].

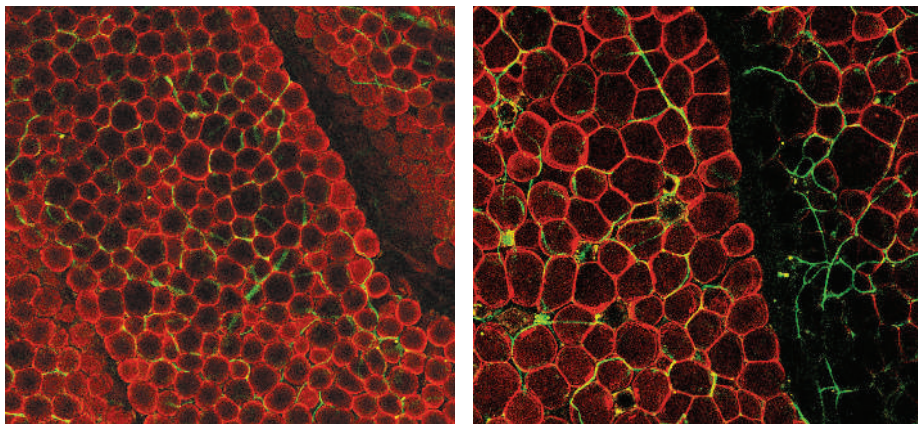
To some researchers, it now seems obvious that fat should control the uptake of glucose, because it may limit further fat accumulation. Lean fat cells produce signals such as adiponectin that promote glucose uptake into tissues and its accumulation as more fat. "We view it as a starvation signal, in essence, signalling that the fat cell is willing to accept additional [stores]," says Scherer. When the fat cells fill up, the signal drops. Cells that become full of fat produce signals such as RBP4 that help stop cells in the body from sucking up glucose and laying down more fat. The glucose instead stays in the bloodstream, where it triggers the problems associated with diabetes.

Inflammatory acts

Engorged fat cells are complicated characters. Over the past decade it has become clear that they send out distress signals, such as TNFα, that can trigger inflammation, and that these contribute to insulin resistance. In their *Science* study, Hotamisligil and his adviser, Bruce Spiegelman of the Dana Farber Cancer Institute in Boston, used a molecule to soak up the TNFα from 'fat' fat cells and found that obese animals got back their insulin sensitivity.

We now know that obese fat cells also produce other inflammatory substances such as the cytokine interleukin-6 (ref. 6). In addition, immune cells called macrophages invade obese fat tissue[7], where they begin contributing to a downward spiral of inflammation. There is overwhelming evidence that obesity and type 2 diabetes are accompanied by chronic, low-level inflammation of fat tissue, says Spiegelman.

Researchers are now trying to work out what causes a 'fat' fat cell, one clogged with lipids, to start secreting a different set of molecules. Hotamisligil's team thinks that the answer may lie in the endoplasmic reticulum (ER) — the cell's centre for folding and processing many proteins. Perhaps because it is crammed with excess lipids and its metabolism is overworked, the ER is unable to properly fold proteins[8].


H. J. LEE & S. SHOELSON

Little and large: fat cells in lean mice (left) send different chemical signals from those in fat mice (right).

This type of cellular stress makes overstuffed fat cells stop making 'healthy' molecules such as adiponectin and begin making 'unhealthy' ones such as RBP4 and TNFα. These trigger insulin resistance in other cells and start the cycle of inflammation, increasing insulin resistance. This chronic inflammation and insulin resistance is now thought to be largely responsible for the plethora of higher disease susceptibilities that obesity brings, including cardiovascular disease, diabetes and cancer.

Fat cells' dynamic new image is changing the way that some researchers think about treating these obesity-related conditions. Treatment usually focuses on losing the fat. But if the problem lies in the adipokines produced by 'fat' fat and the insulin resistance and inflammation they cause, it might be possible to deal with these problems while leaving the fat intact. This approach is needed, some experts say, because of the large number of patients who cannot or will not lose weight.

Reducing RBP4 or raising adiponectin levels in obese patients are two potential treatments for type 2 diabetes. But clinical trials are a long way off because the exact way in which these proteins work is unknown. More likely in the near future is that molecules such as adiponectin and RBP4 may be used as biomarkers in the blood to indicate which overweight patients have higher levels of misbehaving 'fat' fat, and are therefore at risk of diabetes.

Steven Shoelson at Harvard's Joslin Diabetes Center is working on the idea that the chronic inflammation found in the obese causes their diabetes, and he leads a clinical trial to test whether this can be reversed. Starting this year, his team will give half of 120 type 2 diabetics a 14-week course of salsalate, an anti-inflammatory drug similar to aspirin that has been used for more than 150 years. Shoelson was partly inspired to carry out the trial by research published in 1876 — long before the discovery of insulin — suggesting that salsalate could treat diabetics[9].

"We appreciate that perhaps it is not overall obesity that we should look at, but the distribution of fat."
— Philipp Scherer

Hot fat

Treatments that make 'fat' fat cells act like lean ones may be another way to tackle obesity. Diet and exercise do this by starving the body, prompting fat to release its stores into the blood for use in the muscles. But diet and exercise are hard to stick to. And liposuction, which sucks out some of the 'fat' fat cells, is probably not the answer, because the remaining cells are still 'fat' and produce disease-causing molecular signals. So some researchers are trying to exploit the molecular differences between different types of fat cell to convince 'fat' fat cells to reduce their own girth.

In one such effort, Dominique Langin, a clinical biochemist at the Louis Bugnard Institute in Toulouse, France, hopes to exploit the properties of brown fat, a tissue that burns fuel to create heat. Human newborns have a small pad of brown fat on their back, but it quickly disappears. Adults are thought to have only a tiny number of brown fat cells among a mass of white ones.

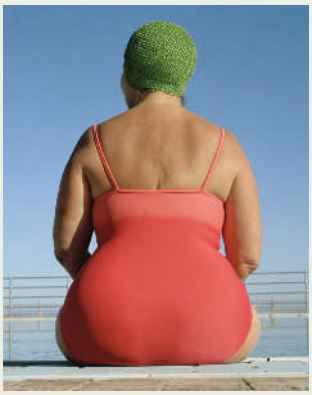
By switching on a gene called *PGC-1α*, Langin's team converted human white fat cells to brown-like ones in the lab dish[10]. The gene seems to switch on others normally active in brown fat. Langin suggests that white fat cells could be converted into brown ones, so that they burn up their fat inside the cell without releasing it into the bloodstream and clogging up arteries. But a drug that can turn on *PGC-1α* in human white fat is still distant, he says.

Hotamisligil hopes to correct aberrant fat cells by relieving the stress on their ER. Last year, his laboratory fed obese and diabetic mice two chemicals that help proteins fold correctly in the ER of fat and liver cells. This improved the animals' ability to control glucose levels[11]. Because both chemicals are already used to treat other diseases, Hotamisligil says he wants to try testing them as a diabetes treatment in the near future.

If this works, it will show that fat cells are not only doing something interesting — as Hotamisligil suspected 14 years ago — but that they can be manipulated for medical use. It takes a lot of willpower to convert a fat body into a thin one. But 'fat' fat cells may be more amenable to a change of character. ■

Kendall Powell is a science writer based in Broomfield, Colorado.

Location, location, location



J. NAUGHTEN/CORBIS

Not all 'fat' fat is created equal: cells in some parts of the body may pump out more of the molecular signals that promote obesity-related disease.

Visceral fat, which wraps around organs and can affect both the lean and obese, seems to be particularly problematic because it dumps signalling molecules directly into blood heading for the liver, the main site where glucose and fat are converted from one to another. Subcutaneous fat, which accumulates under the skin and is hated as cellulite, seems to be less metabolically active and therefore may produce less of these molecules.

These differences might explain a study from 2005 suggesting that weight loss from dieting increased mortality among 3,000 Finns over an 18-year period[12]. The study had the public-health and medical communities in an uproar, as it countered many other surveys showing that weight loss reduces the risks of heart disease and type 2 diabetes, at least over shorter intervals of weeks or months. But however the team did their analyses, "those who maintained their weight fared the best," in the long term, says lead author Thorkild I. A. Sørensen, head of the Institute of Preventive Medicine at Copenhagen University Hospital in Denmark.

Sørensen thinks that otherwise healthy people who lost weight were probably losing subcutaneous fat and beneficial mass from organs and muscles, as well as visceral fat. The higher mortality rates may have occurred because some people lost too much of the good compared to the bad.

Sørensen suggests that moderate dieting and exercise can burn off visceral fat. But sudden and dramatic weight loss might erode subcutaneous fat as well.

"We appreciate more and more that perhaps it is not overall obesity that we should look at, but the distribution of fat," says diabetes expert Philipp Scherer of the University of Texas Southwestern Medical Center in Dallas. "We want to learn how to manipulate different fat depots selectively." **K.P.**

1. Hotamisligil, G. S., Shargill, N. S. & Spiegelman, B. M. *Science* **259,** 87–91 (1993).
2. Abel, E. D. *et al. Nature* **409,** 729–733 (2001).
3. Yang, Q. *et al. Nature* **436,** 356–362 (2005).
4. Graham T. E. *et al. N. Engl. J. Med.* **354,** 2552–2563 (2006).
5. Nawrocki, A. R. *et al. J. Biol. Chem.* **281,** 2654–2660 (2006).
6. Rotter, V., Nagaev, I. & Smith, U. *J. Biol. Chem.* **278,** 45777–45784 (2003).
7. Weisberg, S. P. *et al. J. Clin. Invest.* **112,** 1796–1808 (2003).
8. Özcan, U. *et al. Science* **306,** 457–461 (2004).
9. Ebstein, W. *Berliner Klin. Wochenschr.* **13,** 337–340 (1876).
10. Tiraby, C. *et al. J. Biol. Chem.* **278,** 33370–33376 (2003).
11. Özcan, U. *et al. Science* **313,** 1137–1140 (2006).
12. Sørensen, T. I. A. *et al. PLoS Med.* **2,** e171 (2005).

Treating astrology's claims with all due gravity

SIR — Isaac Newton postulated that there is a force of attraction between any two bodies in the Universe. Your News story 'Gravity passes a little test' (*Nature* **446,** 31–32; 2007) points out that "Isaac Newton's inverse-square law of gravity has given faultless service ever since". I have found a curious way to use Newton's law to draw attention to the difference between science and anti-science for a general audience.

I teach an introductory science class at my university, which typically enrols many non-science majors. During a lecture on the gravitational force, I imply that if planets such as Mars exert a force on any object, including humans, then perhaps there is something to astrology's idea that celestial bodies exert a force of influence on our lives. I encourage my students to undertake a test I have designed for this notion.

I present the students with 12 randomly numbered horoscopes from the previous day, with the corresponding signs of the zodiac removed. I ask each student to record the horoscope that best describes the day she or he had, and the astrological sign (for example, Aries) corresponding to her/his birthday. My scientific hypothesis is that planets may exert a force on our bodies, but it is purely random — 1 out of 12 (8.3%) — whether a horoscope foretells the events of one's life.

I am pleased to report that, as Shawn Carlson has noted, "astrology failed to perform at a level better than chance" (*Nature* **318,** 419–425; 1985). The results from my classes are: 8.0% ($n = 163$ students), 8.4% ($n = 155$), 7.0% ($n = 143$), 8.0% ($n = 138$) and 8.0% ($n = 100$). In other words, as John Maddox has commented "astrology is a pack of lies … There is no evidence that the positions of the planets can affect human behaviour" (*Nature* **368,** 185; 1994).

I encourage science teachers to try this approach when they are presented with an opportunity, as this exercise inspires genuine scientific inquiry. For example, students have countered that a certain astrologer may not be qualified to read the stars. I have addressed this question by using horoscopes from different newspapers (*Chicago Tribune, Columbus Dispatch, Los Angeles Times* and *Washington Post*) and online sources (Astrology.com, DailyHoroscopes.com). Other students have noted that small groups, with only two students of a particular sign, may obtain a result that is significantly greater than 8.3%. This presents an opportunity to discuss the value of an adequate sample size.

Finally, it is worth reporting that my students are so engaged by this exercise that they actually want to use Newton's law of universal gravitation to calculate force values. In case you are curious, Mars, at its present distance of 264 million kilometres from Earth, is exerting a force of approximately 50 nanonewtons on your being.

Steven K. Lower
The Ohio State University, 275 Mendenhall Laboratory, Columbus, Ohio 43210, USA

Conflict and cancer research in Arizona

SIR — Your News Feature 'The Arizona experiment' (*Nature* **446,** 968–970; 2007) mentions that Robert Pettit "lost" directorship of the Cancer Research Institute at Arizona State University, as though he were an absent-minded professor who had just misplaced it. In fact, after disagreements with the university's president Michael Crow, Pettit was removed from the position and the institute was effectively closed down. All the personnel were transferred to the university's Biodesign Institute. After a year of problems and disputes, which included measures taken to stop them obtaining any new funding, the group was terminated. The reasons given by Crow and the Biodesign Institute's head George Poste, according to articles that appeared in local newspapers, were that the research was out of date and that the group had not kept up with recent advances.

A less euphoric assessment of Crow's administration and its effect on morale can be found in an article by Megan Irwin in the *Phoenix New Times* (www.phoenixnewtimes.com/2007-04-26/news/asu-inc).

John C. Knight
13646 East Westland Road, Scottsdale, Arizona 85262-5850, USA

Pathologists needed to cope with mutant mice

SIR — The number of genetically engineered mouse mutants is rising substantially, as highlighted in your Editorial 'Mutant mice galore' (*Nature* **446,** 469–470; 2007). However, there is insufficient manpower and expertise in comparative pathology to characterize and validate these model animals effectively. We ask the scientific community to recognize this crisis and help address the issue before it becomes catastrophic.

The International Mouse Knockout Consortium is creating a critical, but unmet, need for expert comparative pathologists knowledgeable in mouse biology and human disease. Effective mouse pathology requires an understanding of mouse biology and a knowledge base that is not possessed by most investigators or pathologists. Funding agencies recognize the problem, but are not sufficiently addressing it, and the funding mechanisms of the US National Institutes of Health do not allow for training in pathology. Where are the mouse pathologists of the future, and who is going to train them?

Scientists involved in creating the huge mutant-mouse population need to recognize this need and help find a solution. The governments funding the mice and their databases are failing to create the necessary human resources. One potential solution could be a partnership among academic institutions, industry and government to develop and support an 'electronic consortium' of existing mouse pathology specialists. Their collective wisdom could be shared with interested young pathologists, using distance-learning tools.

Robert D. Cardiff
Center for Comparative Medicine, University of California at Davis, Davis, California 95616, USA

Open journals' records to give reviewers their due

SIR — Sydney Brenner and Richard Robert's request in Correspondence (*Nature* **446,** 725; 2007) for authors to conserve records of their work and make them freely accessible is of great importance to historians of science.

However, unlike an artist's preparatory sketches or a novelist's drafts, scientific papers describing major discoveries have gone through the process of peer review. Reviewers often make significant contributions in shaping discoveries. They suggest new experiments, propose novel interpretations and reject some papers outright. Clearly, this is also important 'behind the scenes' work by scientists usually at the forefront of their discipline, and is an intrinsic part of the scientific process. It is well worth keeping a record of such work, for no history of science will be complete and accurate without it.

I therefore propose that journals' records should be made publicly available after an adequate lapse of time, including the names of reviewers and the confidential comments exchanged between editors and reviewers. The Nobel Foundation makes all its records available after 50 years, as do many governmental and other institutions. This delay may be reduced for scientific journals to, perhaps, 15 or 20 years. This is also likely to have a positive impact on the peer-review process itself.

The scientific community and future historians will gain from this transparency and from full knowledge of all the events that have contributed to a great discovery.

Ariberto Fassati
Wohl Virion Centre and MRC–UCL Centre for Medical Molecular Virology, Division of Infection and Immunity, University College London Medical School, 46 Cleveland Street, London W1T 4JF, UK

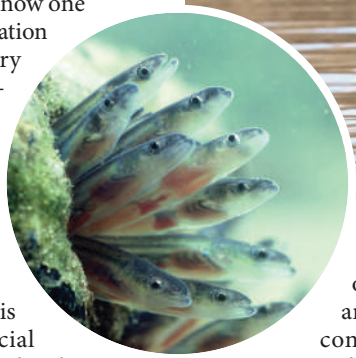
COMMENTARY

Can bans stimulate wildlife trade?

Proactive management of trade in endangered wildlife makes more sense than last-minute bans that can themselves increase trading activity, argue **Philippe Rivalan** and his co-authors.

Wildlife trade threatens around one-third of birds and mammals worldwide[1]. A further 1,000 timber species are threatened with extinction due to felling[2], and 75% of fisheries are fully or over-exploited[3]. Yet the use of natural resources directly supports the livelihoods of around 200 million people, including many of the poorest on Earth[4]. Clearly, sustainable management of these resources is vital to the well being of both human and ecological communities. But can we meet the twin challenges of poverty reduction and species conservation in the twenty-first century?

Since 1975 one of the most effective international treaties seeking to conserve biodiversity has been the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). CITES is now one of the best supported conservation agreements, with 171 signatory parties. During the next Conference of the Parties (CoP), to be held 3–15 June 2007 in The Hague, the Netherlands, discussions will consider a range of controversial proposals to regulate wildlife trade. The most severe restriction that CITES can enforce is an explicit ban on commercial trade of wild species threatened with extinction. We report here concerns that such bans can themselves lead to an increase in trade of vulnerable species, and suggest ways to improve species management that would be in the interests of traders and conservationists alike.



J. E. ROCHE/NATUREPL.COM

R. JACKMAN/PHOTOLIBRARY.COM

The European eel may suffer further heavy decline if CITES delays its listing.

CITES provides an international legal framework to regulate trading of animal and plant species, including some trophy hunting and the ivory trade. Of the around 33,600 species protected by CITES, roughly 800 are listed under Appendix I to the Convention because they are threatened with extinction and so are protected by wildlife trade bans. Another 32,500 species are listed under Appendix II because they may become threatened with extinction unless CITES regulates trade through permits and licences. The remaining 300 species are listed on Appendix III by countries seeking assistance to control trade in that species. Some have characterized CITES as an old-fashioned command-and-control convention[5], in contrast with current trends towards market-based and incentive-driven conservation[6]. Critics of CITES argue that it has sometimes failed to effectively regulate trade and to enforce bans[5].

Value added

Concerns were first raised in 1985 that 'uplisting' species to a more restrictive appendix could make them more valuable to traders and consumers, but there has been no wide-ranging analysis to support this idea. Here, we analyse legal imports of wild-collected specimens from animal species uplisted from Appendix II to Appendix I between 1980 and 2003. Uplisting generally takes effect between 240 and 420 days after submission of the proposal, leaving a year for traders to acquire specimens or to clear stocks before trade restrictions apply. Our analysis shows that legal volumes of imports increased during this transition period (see graph, overleaf).

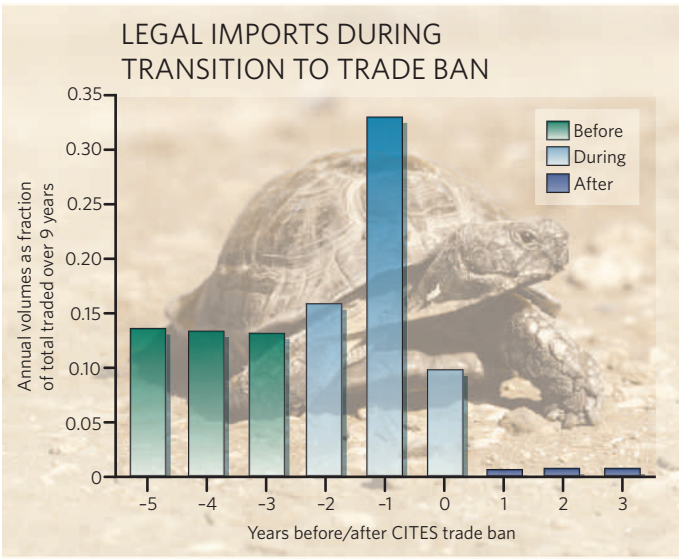
For this analysis, trade data for each species were assigned to three time periods of 3 years each: before the uplisting proposal was submitted; during the uplisting evaluation; and after the uplisting came into force. We used reported data on gross import volumes of legal trade in wild specimens from the CITES database (www.cites.org/eng/resources/trade.shtml), and only considered species with more than five specimens traded over the whole period, for a total of 46 animal species.

The graph overleaf shows that trade volumes declined significantly after the commercial ban was in force, but a peak in trade was detected 1 year before the ban started. This peak corresponded, on average, to an increase of 135% in the trade volumes compared with previous years (see supplementary information for details of method).

Although further field-based research is needed to assess how such spikes in trade might affect individual species, in some cases the reported trade volumes alone are of concern. For example, the peak volumes of 2,800 Kleinmann's tortoises and 5,500 Geoffroyi cats imported during the transition period,

represent one-half and one-tenth, respectively, of their total estimated mature population sizes[7,8]. In addition to increases in legal trade, uplisting of species to Appendix I may also lead to an increase in their commercial value and in illegal trading. For example, the price of rhino horn on Korean markets increased by more than 400% within 2 years of their uplisting, which in turn coincided with a sharp increase in poaching of black rhinos and in illegal trade in rhino horn. Any such illegal trade is not picked up in CITES-reported data once the uplisting takes effect and so it is much harder to track.

All available information, including that intended to protect species against trade, can be used by wildlife traders to make commercial decisions. When CITES parties meet in June to discuss 36 proposals to change species listings, they should consider that traders may anticipate uplisting during the transition period.



A. F. PAPAZIAN/NHPA

Early intervention

At the very least, our findings suggest that CITES authorities will need to use extra vigilance in controlling permits during transition periods and in adhering to quotas. One possibility — reducing the time between the proposed uplisting and the ban coming into force — is not practical because parties require time to evaluate uplisting proposals, and the deadlines are specified in the Convention text and so are difficult to amend. A better solution may be to find ways to manage Appendix II species so as to reduce uplistings to Appendix I. CITES might also have more chance of success if commercially important species were included in Appendix II at a point when there is still time for management to improve conservation.

This proactive approach could work well with commercially important species associated with unmanaged and illegal trade that, historically, parties have been slow to list under CITES. In the past decade, CITES has become more involved in helping to manage such high-value commodity species, including sturgeon for caviar, the sought-after humphead wrasse, and mahogany and ramin hardwood trees for timber. Listing of these species has increased collaboration with mainstream producer organizations such as the Fisheries Department of the UN Food and Agriculture Organization (FAO) and the International Tropical Timber Organization. But changes have been slow. The listing of big-leaf mahogany on Appendix II was mooted 12 years before the proposal was accepted. During that time deforestation reduced the area of Latin American forests by more than a tenth[9]. Similarly, CITES and the FAO took 5 years to agree to collaborate on managing trade of CITES-listed commercial fishery species.

"Can we meet the twin challenges of poverty reduction and species conservation in the twenty-first century?"

Timely listings will be particularly important at the coming CoP. Of the species proposed for listing in Appendix II, seven are commercially important: three timber and four fish species. The timber species include Central American brazilwood, rosewoods and a Latin American cedar. The fish species include the spiny dogfish, porbeagle and European eel, all of which have declined in Northern Hemisphere waters.

Sustainable trade

For some of these species, a lengthy delay in listing could have important conservation consequences. For example, European eel stocks are estimated to have declined to between 9% and 19% of their baseline abundance since 1980 and much of the annual population production is now commercially traded[10]. Meanwhile, trade volumes have halved over the past decade and the value of exports has increased tenfold. A prolonged delay in listing this species and coordinating management action could result in heavy population declines.

Seven more proposals call for increases in protection through Appendix I listing, but at this CoP, only two are for uplisting from Appendix II. Of these, uplisting of the Guatemalan bearded lizard risks the dangers shown in the graph, because only 170 to 250 individuals are estimated to remain in the wild. The other uplisting proposal concerns the Asian slow lorises, for which it is not clear whether or not levels of trade and population reduction warrant inclusion in Appendix I.

Over the years, CITES has diversified its activities by encouraging positive measures to facilitate sustainable trade that will not negatively affect species survival. These measures include downlisting of species from Appendix I for ranching purposes and introducing trophy-hunting quotas for Appendix I-listed species[11]. In line with this flexibility, a proposal for the current CoP seeks to ensure that CITES listing benefits species conservation and rural livelihoods. In addition, there is some interest in promoting certification of CITES trade, to improve local management in biodiversity-rich but economically poor nations, and to share with consumers the enforcement costs currently borne by governments.

In our view, CITES can help to deliver the sustainable use of natural resources as required by the overarching Convention on Biological Diversity. The international community needs to find ways to contribute effectively to the UN 2010 Biodiversity Target and the 2015 Millennium Development Goals. CITES faces this challenge from a sound basis of experience, expertise and infrastructure, and an ability to evolve. As they make decisions on the proposed listings and livelihood issues, CITES parties will demonstrate whether or not they have the resolve to meet the ongoing challenges of poverty reduction and species conservation. ■

Philippe Rivalan, Virginie Delmas, Elena Angulo, Leigh S. Bull, Richard J. Hall and Franck Courchamp are at University Paris-Sud, Laboratoire Ecologie, Systématique et Evolution, UMR CNRS 8079, Orsay F-91405, France.
Alison M. Rosser and Nigel Leader-Williams are at the Durrell Institute of Conservation and Ecology, University of Kent, Canterbury, CT2 7NR, UK.

1. Baillie, J. E. M., Hilton-Taylor, C. & Stuart, S. N. *2004 IUCN Red List of Threatened Species* (IUCN, Gland, 2004).
2. Oldfield, S. F., Lusty, C. & Mackinven, A. *The World List of Threatened Trees* (World Conservation Press, Cambridge, 1998).
3. FAO Fisheries Department *State of the World's Fisheries and Aquaculture 2004* (FAO, Rome, 2004).
4. Van Rijsoort, J. *Non Timber Forest Products* (National Reference Centre for Nature Management, Wageningen, 2000).
5. Hutton, J. M. & Dickson, B. *Endangered Species, Threatened Convention: the Past, Present and Future of CITES* (Earthscan, London, 2000).
6. Hutton, J. & Leader-Williams, N. *Oryx* **37,** 215–226 (2003).
7. Nowell, K. & Jackson, P. (eds) *Wild Cats: Status Survey and Conservation Action Plan* (IUCN, Gland, 1996).
8. Perälä, J. in *2006 IUCN Red List of Threatened Species* (IUCN, 2003).
9. FAO Forest Resources Assessment (FAO, Rome, 2001).
10. FAO Fisheries Report No. 833; COP14 Doc. 68 Annex 3 (FAO, Rome, 2007).
11. IUCN *Trade Measures in Multilateral Environmental Agreements* (Submitted to The Economics, Trade and Environment Unit, UNEP, 2000).

Supplementary Information accompanies the paper at www.nature.com/nature.

BOOKS & ARTS

Time for a change?

Two books on the AIDS pandemic in Africa challenge assumptions at the heart of the UN's response.

The AIDS Pandemic: The Collision of Epidemiology with Political Correctness
by James Chin
Radcliffe Publishing: 2007. 336 pp. £27.50, $39.95
The Invisible Cure: AIDS in Africa
by Helen Epstein
Farrar, Straus & Giroux: 2007. 336 pp. $25

H. BURDITT/REUTERS

Positive action: marchers in Cape Town campaign to prevent 2 million new HIV infections by 2010.

Stephen Lewis and Paula Donovan

Has the tide turned away from AIDS orthodoxy? Two books — strikingly different in tone and character — recount the global response to the AIDS pandemic with words of recrimination for the United Nations (UN). Helen Epstein and James Chin each raise searching questions about methods and motives, saving their most pointed barbs for UNAIDS, the body that coordinates the work of ten separate organizations within the UN system on the pandemic and tracks its spread.

There is a liberating quality to the way in which the arguments are raised and, if these authors are right, a major overhaul of the international AIDS response is overdue.

Chin, a professor of clinical epidemiology at the University of California, Berkeley, has an axe to grind. He has long felt that the work of measuring HIV/AIDS and projecting its course is in the hands of scientists who lack epidemiological know-how. The results are inflated statistics and predictions of a global Armageddon with no basis in fact.

Over several pages, Chin describes how misapplied mathematical models have churned out exaggerated numbers, all flowing from the faulty assumption that national prevalence rates can be estimated by testing women in urban prenatal clinics and extrapolating to general populations. On this score, Chin has been vindicated by the US-funded Demographic and Health Surveys — the more accurate data gathered from randomly selected households in cities and rural areas — that are forcing UNAIDS to lower its estimates of national prevalence rates dramatically in country after country. (UNAIDS simply attributes these reductions to improved surveillance and will, says Chin, "ride to glory" on the myth that further declines prove the success of its prevention programmes, particularly among the young. In fact, he says, because most AIDS epidemics peaked in the mid-1990s, rates are receding naturally.)

Having been right on prevalence, Chin has to be taken seriously when he dismisses UNAIDS' doomsday predictions for India and China. Yes, this pandemic is an unparalleled public-health emergency, he says, but the double-digit prevalence rates seen in parts of Africa will never be reached elsewhere. That's because HIV takes hold among members of high-risk groups (men who have sex with men, injecting drug users, female sex workers), but only fans out to general populations where patterns of multiple, concurrent sexual relationships are the norm. Thus in many African countries, HIV travelled from a member of a high-risk group to each of his or her regular sex partners, then to their multiple regular partners, who introduced the virus to other networks until a vast web formed. It was not the number of sex partners — people everywhere average roughly the same number over their lifetimes — but the synchronicity of the encounters. According to Chin, in most of the world outside east and southern Africa, a culture of serial rather than concurrent sex partners has so far, and will continue to, confine HIV epidemics to high-risk groups. He predicts a slow rise in prevalence rates due primarily to life-prolonging treatment and to new infections, which from here on, he says, will be confined largely to people whose regular sex partners are HIV-positive.

If Chin is right — and to the non-scientist, at any rate, his textbook-like narrative is certainly provocative — the UN has wilfully deceived. Why? Because, he contends, the larger the numbers, the greater the carnage and the more the money flows in. Because associating the continent's horrific AIDS ordeal with African sexual practices might seem like racial stereotyping. Because it's easier to conduct prevention programmes among the general public than among the high-risk groups at society's edges. But why not over-reach, spreading prevention messages to entire populations, including those at highest risk? Because, Chin argues, given limited resources, only targeted prevention programmes can protect those most likely to become infected. Yet in place of that simple piece of logic, he sees an AIDS *gestalt* created in the service of fundraising.

UNAIDS identifies poverty, gender inequality, discrimination and lack of access to healthcare as underlying causes of sub-Saharan Africa's plague. Chin asserts that these problems must be addressed "because they create major barriers to effective HIV prevention and treatment programs, but they are not the

primary or even the major determinants of high HIV prevalence". We're not sure that social scientists would agree (the line from poverty and gender inequality to high-risk commercial sex work seems fairly direct), but here again, it's hard to avoid the logic in the numbers.

There's much in Chin's book that Helen Epstein, author of *The Invisible Cure: AIDS in Africa*, would find palatable. Like Chin, she holds the bold view that the virus spreads beyond high-risk groups to the general population only through web-like networks of concurrent sex partners.

With elegant prose, a scientific background and a journalist's searching anecdotal eye, Epstein combines personal research and corroborative evidence from others to posit the view that where Africa's AIDS rates are highest, the key difference is not the numbers of sexual partners, but the timing. She then applies her theory to Uganda, the one country in Africa where a culture of concurrent sex partners was well entrenched and yet the prevalence rate of HIV/AIDS has been reduced dramatically since 1990.

Epstein describes how President Yoweri Museveni rallied his country in the 1980s with the mantra 'Zero Grazing' that cautioned Ugandans not to have more than one partner at a time. She acknowledges the application of 'ABC' — abstain, be faithful and use condoms — but asserts that an abrupt end to the practice of concurrent relationships was the decisive factor in reducing prevalence.

How was it brought about? Here we see Epstein at her best, explaining why Uganda succeeded where others failed. Above and beyond the messages from government bureaucrats, and a world away from donor-driven aid 'packages', the "personalized, informal, intimate, contingent, reciprocal nature of African society" led Ugandans to draw their own conclusions and fashion their own grassroots defence. Concludes Epstein, "the open discussions led by government field-workers in small groups of women and churchgoers, the compassionate work of the home-based care volunteers, the courage and strength of the women's-rights activists helped people see AIDS not as a disease spread by 'others' but as a shared calamity, and this made discussion of sexual behavior possible without seeming preachy, condescending or out of touch".

Wealthier countries, such as South Africa and Botswana, looked to imported commodities and slick advertising campaigns, but in Uganda, the shocking pervasiveness of death was mellowed by the traditional African principle of *ubuntu*, or shared humanity, resulting in an indigenous response that stirred the collective conscience. Epstein speaks of the neighbourly exchange of caregiving as though it was a social movement, and in the mind's eye, the reader is convinced.

And then, like Chin, Epstein goes after UNAIDS. Mind you, her criticism is almost Victorian in its gentility, albeit unmistakable in its target. Where Chin uses a hatchet, Epstein wields the scalpel. She argues persuasively that the UN has long known that reduction in the number of sexual partners has been a factor wherever rates have fallen, from Uganda to San Francisco, and yet it refuses to act on it. Is it lack of respect for indigenous cultural awareness and survival instincts? She recounts how in 1993, a statistician now in the top ranks of UNAIDS misrepresented (mistakenly, it seems) findings about Uganda's success, erroneously claiming that researchers had noted an uptake in condom use and delay in sexual initiation, but no significant reduction in concurrent partners. The distortion prevailed. "It was only in 2006 that UNAIDS officials began to stress that the reduction of multiple sexual partnerships should be a key goal for AIDS prevention programmes in southern Africa," she reports. When the organization's executive director Peter Piot was asked about the omission in a list of questions Epstein sent him in 2004, he answered every question but that one.

The issue that lurks at the back of the mind of the reader in the case of both Chin's and Epstein's arguments is, what next? If UNAIDS has been locked in a rut of culturally questionable and epidemiologically flawed approaches, can it be rehabilitated? Chin would have us look harder at the scientific facts, and then start afresh with a new set of assumptions and projections; he would eliminate conflicts of interest by segregating epidemiologists from the officials concerned with advocacy and fundraising. Epstein would add to that the plea that programme planners rediscover the indigenous wisdom of African culture, which has successfully withstood threats since the dawn of humankind. Both books are guaranteed to spark animated discussion. Together, they pose the first open challenge to the UN's role in the most eviscerating plague in human history. ■

Stephen Lewis is former UN special envoy for AIDS in Africa, and Paula Donovan is an international HIV/AIDS consultant. They are in the process of setting up an international AIDS advocacy organization, AIDS-Free World.

More on AIDS in Africa

When Bodies Remember: Experiences and Politics of AIDS in South Africa
by Didier Fassin (University of California Press, 2007).
28: Stories of AIDS in Africa
by Stephanie Nolan (Portobello Books, 2007).

Winged beauty

Moths, generally active at night, are often less colourful than butterflies. But there are exceptions. The brilliant colours of the *Pseudodirphia menander* (pictured) are not ones that attract attention from afar in its natural habitat — in the forest undercanopy many leaves have pink and red hues. Closer up, the red colour might act as a warning. *100 Butterflies and Moths* (Belknap Press) presents portraits of some of the 10,000 species estimated to inhabit the Guanacaste conservation area in northwestern Costa Rica. The photographs, taken by Jeffery Miller, and the accompanying, informative descriptions could serve as a field guide for a visit to the Costa Rican forest or be equally appealing to the armchair naturalist.

J. C. MILLER, D. H. JANZEN, W. HALLWACHS

N. SIMPSON/EMPICS SPORT/PA PHOTOS

Football crazy: can Chelsea football supporters be seen as "bees in a hive" in terms of darwinian evolution?

Selling evolution

Evolution for Everyone: How Darwin's Theory Can Change the Way We Think About Our Lives
by David Sloan Wilson
Delacorte Press: 2007. 400 pp. $24

Mark Pagel

Evolutionary biologists — those enthusiastic foot-soldiers of Darwin's grand notion that life evolves by a process of descent with modification — cannot understand why so many people reject the great man's theory, and often in favour of some form of creationist account of the existence and diversity of life on Earth. In the opening pages of David Sloan Wilson's new popular-science book, hopefully entitled *Evolution for Everyone*, we discover that 54% of adults in the United States prefer to believe that humans did not evolve from some earlier species. What makes this figure surprising is that it is up from 46% in 1994.

Where have the evolutionists gone wrong? One answer is staring them in the face but not often noticed. A double irony is that it derives from their own theory: if people differ in the strength or conviction of their religious beliefs, if children tend to acquire religious beliefs from their parents, and if religious people, for whatever set of reasons, tend to have more children, then it follows from Darwin's theory of evolution by natural selection that religious belief will spread: *quod erat demonstrandum*.

The other answer is the one that evolutionary biologists, including Sloan Wilson, prefer and it provides him with the impetus for this agreeable little book: that if the evidence for darwinian evolution is presented clearly enough and often enough, any reasonable person will come around to the darwinian view. What is there to say? The usual answer, that we share more than 98% of our genes with chimpanzees, is becoming hackneyed. It is the strangeness of human behaviour that really puts the darwinian view to the test. And here there is much to discuss. We have enormous brains that make us shrewd beyond belief in comparison to other animals, we have the only fully developed symbolic language on the planet, we cooperate with and engage in elaborate task-sharing and reciprocal relations with people we don't know, we help the elderly, give money to charities, put on matching silly shirts to attend football matches, obediently wait in queues, die for our countries or even sometimes for an idea, and we positively ripple and snort with righteousness and indignation when we think others don't do some of these things. We even have a word for this sense of how others ought to behave — morality. Chimpanzees, and for that matter other animals, aren't like this. No wonder the creationists don't believe the darwinian account.

A popular view among students of human evolution is that special ideas may be needed to explain what is sometimes called our 'extreme sociality' — the helping, reciprocity and morality. Sloan Wilson is among the principal advocates of the view that humans have evolved by a process of 'group selection' in which groups of people — our hunter-gatherer or early tribal ancestors — worked together in ways that allowed them to outcompete other groups. Over time, this process moulded our psychology and social behaviour so that we became, as Sloan Wilson puts it, like cells in a body, or bees in a hive, devoted to the well-being of our group. Laughter, music, dance and religion are interpreted as aids to promoting a sense of group membership and mutual well-being. Sloan Wilson pays particular attention to ways in which religions prohibit murder and other antisocial behaviour within the group but offer rewards for those who use it against people outside the group.

The group-selection account is seductive, explanatory and may even be right, but what about our tendencies to cheat, deceive, manipulate and coerce? Why do we need so many laws, police forces, jails, speed cameras and tax offices? Why do we gossip incessantly about others' behaviours and reputations? Why do we compete so strenuously to get ahead and pay so much to get our children educated? Cells in our bodies and bees in hives are much better behaved, and don't have big brains like we do. Might it just be that the 3.5 billion years of selfish darwinian natural selection that preceded the invention of humans bequeathed us a legacy — of a species whose behaviours are largely driven by the selfish desire to promote ourselves and our offspring? Could it be that we humans acquired our supreme intelligence at least in part to manipulate social systems in ways that promote our individual reproductive success?

The intellectual ravine separating these two camps rises on the one side to a view of human behaviour as being for the good of the group and on the other to the view that it makes use of the group for individual benefit. It is a delicate and subtle debate and Sloan Wilson's popular accounts in *Evolution for Everyone* make for enjoyable and thoughtful reading. But perhaps even Sloan Wilson should not expect to change people's minds about religion. If our minds evolved to help us wade through the complexity of social life, to use groups for our own gain, and to help us rebound from 'the slings and arrows of outrageous fortune', which set of beliefs, on balance, will be more useful, religious ones (whether true or not) or a belief in natural selection? ■

Mark Pagel is professor of evolutionary biology at the University of Reading, Reading RG6 6AH, UK.

NEWS & VIEWS

ASTROPHYSICS

Water worlds in the making

Roy van Boekel

Meticulous observations of the disk of gas and dust around one young star seem to imply icy, comet-like bodies in the disk's inner regions. Could these be the building-blocks of water-rich planets like Earth?

The disks of dust and gas around newborn stars provide a natural laboratory for studying the formation of planetary systems. But even the nearest young stars are so far away that conventional telescopes lack the sharpness of vision to see into the very centre of the disk, where habitable planets such as Earth could form. Using the Keck Interferometer — a system of two telescopes on the summit of Mauna Kea, Hawaii — J. A. Eisner has zoomed in on this 'terrestrial region' around one young star. The results[1] can be found on page 562. He measures the disk geometry on scales of less than one astronomical unit (AU) — the distance from Earth to the Sun — and directly locates hydrogen gas plunging onto the star. He finds water, too, in the inner disk, arguably stemming from evaporating icy bodies that might even now be forming water-rich, Earth-like planets.

The disks around stars consist mostly of hydrogen gas, but they also contain molecules such as carbon monoxide (CO) and water that are much less abundant, but easier to detect by virtue of their much stronger spectral signatures. About 1% of the disk mass consists of tiny particles of solid material. Put enough of this 'dust' together, and you can make a rocky planet like Earth, or the core of one of the giant gas planets found farther out in our Solar System.

Of the more than 200 'exoplanets' now known to be orbiting stars other than the Sun, a preponderance are giant gas planets close to their parent star. That is partly because these planets are easiest (or rather, least difficult) to detect. Even so, it seems that our Solar System's architecture, which is apparently so favourable for the development of life — rocky planets close to the star, gas giants farther away, all on approximately circular orbits — is not common[2]. But exactly how uncommon is it?

To find out, we need to go back to the planetary cradle. In most of a circumstellar disk, gas and dust are mixed. In the region of interest close to the star, however, the temperature is so high that solid material vaporizes, and hot gas swirls its way towards the star. Dynamical interactions between this hot gas and forming planets, which cause planets to migrate through the disk, are considered crucial in determining the final architecture of a planetary system[3].

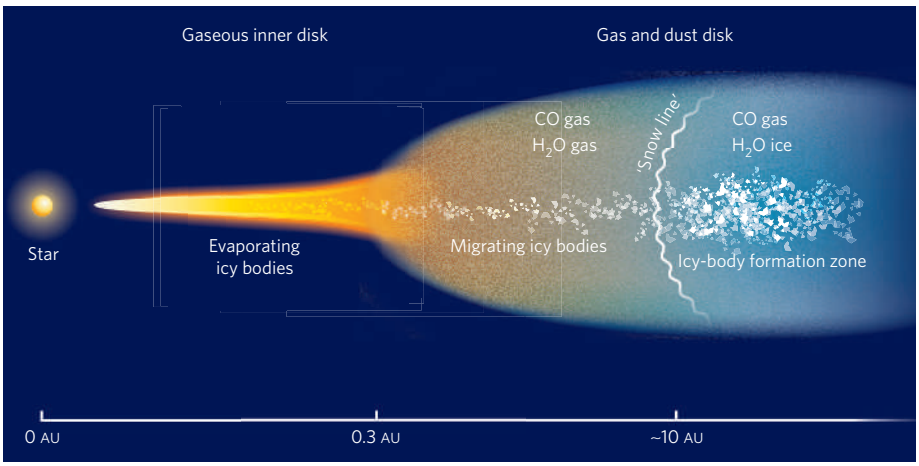


Figure 1 | Ice advance. A planar 'disk' of gas (mainly hydrogen, but with observable components of water and carbon monoxide, CO) and dust surrounds a young star (diagram not to scale). The farther away from the star, the lower the temperature. For a star of the luminosity of MWC 480, only gas is present within about 0.3 AU of the star (1 AU is the Earth–Sun distance); at larger radii, there is a mixture of gas and dust. Icy, solid bodies will form beyond a 'snow line' about 10 AU distant; closer in, icy bodies will evaporate. Eisner[1] detects water emission in the inner disk around MWC 480, but accompanying CO emission is not seen. These observations support the idea that icy solid bodies are migrating into the innermost disk from beyond the snow line — where, although water is solid, CO is still gaseous and is therefore not incorporated in solid bodies. Earth's water-rich composition could reflect the presence of such objects among the bodies that merged to form it.

The spatial scales involved in studying the inner-disk region present a formidable challenge. At the distance of the nearest circumstellar disks, the angular separation between the Sun and Earth would seem about 0.000003° to us — the thickness of a human hair seen from 2 kilometres away. No telescope offers that kind of resolution, and so astronomers have had to rely on indirect methods to study the disk material in the terrestrial region. They analyse emission lines from gaseous material at very high wavelength resolution, but low spatial resolution, allowing reconstruction of velocities in the gas through the Doppler shifts in their wavelengths. Using Kepler's laws of orbital motion, these velocities can be translated into distance from the star. This way, gaseous inner disks have been detected with CO emission very close to the star, at less than 0.1 AU, and hot water vapour at slightly larger radii of less than 0.3 AU (ref. 4).

Eisner[1] uses the two 10-metre-aperture Keck telescopes in symbiotic mode — making them work together as one giant 85-metre eye — to zoom in on the innermost disk region of a young star named MWC 480. He finds that the inner edge of the dusty disk is about 0.3 AU from the star, with additional gas closer in. There is evidence for hot water vapour at around 0.15 AU, confirming earlier findings around other stars[4]. On the smallest scales (<0.1 AU), hot hydrogen gas is accreting onto the star. These observations agree with earlier models designed to explain spatially unresolved spectra[5], but contrast with recent spatially resolved observations of another young star, in which hydrogen emission was shown to arise mostly at much larger radii[6].

Perhaps even more interesting than what Eisner saw is what he didn't see: emission from CO. In sources surveyed so far[7], hot water emission is accompanied by usually stronger CO emission. How did water get to MWC 480's inner disk without bringing CO with it, as would be expected with a steady

accretion flow? There is one viable mechanism (Fig. 1): that, hidden from view deep within the disk, comet-like bodies rich in water ice are sent towards the inner disk through gravitational interactions. In the inner disk, the water is released as the icy bodies evaporate. If these bodies form at the temperatures of 20–150 kelvin found in almost the entire disk of MWC 480, water will be ice, but CO will still be gaseous. Such bodies will therefore contain lots of water, but little or no CO.

The idea that icy bodies are on the move in the innermost regions of a solar system in the making would confirm theories of Earth's own formation. Earth is thought to have been produced through the merging of large numbers of smaller bodies, some of which came from larger distances and contained water ice, thus supplying large quantities of water: Earth's interior is estimated to contain between one and ten times more water than its oceans[8]. Where Earth's surface water came from is, incidentally, a matter of debate: rival theories hold that the oceans were 'sweated' by the young planet through volcanic eruptions, or that they were the product of comet-like bodies that collided with the planet after it had formed[9,10]. Unfortunately, MWC 480 is much younger than our Solar System was when Earth's formation was complete, and so Eisner's observations do not provide any new handle on this controversy.

One factor speaks against the interpretation of Eisner's results as the signature of comet-like icy bodies. This is the reasonably high rate at which material is falling onto MWC 480's central star, which makes an intermittent, cometary origin of the accreting material less likely than a steady, gaseous accretion flow. Such steady accretion would be more in line with existing ideas and observations of other stars, but puts us back to square one with the problem of why we see water, but no CO. High-resolution spectroscopy to confirm the detection of hot water vapour and to put stringent limits on the CO emission in the MWC 480 system are urgently needed as the next steps towards solving the puzzle. ■

Roy van Boekel is at the Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany.
e-mail: boekel@mpia-hd.mpg.de

1. Eisner, J. A. *Nature* **447,** 562–564 (2007).
2. Udry, S., Fischer, D. & Queloz, D. in *Protostars and Planets V* (eds Reipurth, B., Jewitt, D. & Keil, K.) 685–700 (Univ. Arizona Press, Tucson, 2007).
3. Papaloizou, J. C. B. *et al.* in *Protostars and Planets V* (eds Reipurth, B., Jewitt, D. & Keil, K.) 655–668 (Univ. Arizona Press, Tucson, 2007).
4. Carr, J. S., Tokunaga, A. T. & Najita, J. *Astrophys. J.* **603,** 213–220 (2004).
5. Muzerolle, J., Calvet, N. & Hartmann, L. *Astrophys. J.* **55,** 944–961 (2001).
6. Tatulli, E. *et al. Astron. Astrophys.* **464,** 55–58 (2007).
7. Najita, J. R. *et al.* in *Protostars and Planets V* (eds Reipurth, B., Jewitt, D. & Keil, K.) 507–522 (Univ. Arizona Press, Tucson, 2007).
8. Lecuyer, C., Gillet, P. & Robert, F. *Chem. Geol.* **145,** 249–261 (1998).
9. Morbidelli, A. *et al. Meteor. Planet. Sci.* **35,** 1309–1320 (2001).
10. Mumma, M. J. *et al. Science* **292,** 1334–1339 (2001).

AGEING

When less is more

Adam Antebi

Restricting dietary intake is one way to promote longevity. The identification of two genes that specifically mediate this effect in worms provides insight into the molecular mechanisms underlying ageing.

Dietary restriction — a reduction of food intake by 40–60% without malnutrition — has remarkable benefits for health and lifespan, extending the survival of species as diverse as yeast, worms, flies, rodents and perhaps even primates. Yet despite intensive study, the molecular basis of the effects of dietary restriction in animals has remained largely elusive. Elsewhere in this issue, two groups[1,2] report a role for a pair of evolutionarily conserved proteins — PHA-4 and SKN-1 — in conferring extended survival under dietary restriction in the small roundworm *Caenorhabditis elegans*. Both PHA-4 and SKN-1 are transcription factors, regulating the expression of many genes. Moreover, evidence suggests that they may also trigger hormones that coordinate physiological responses to dietary restriction.

Restricting dietary intake in worms can be achieved by diluting their bacterial food. At optimum conditions for dietary-restriction-induced longevity, worms typically live 20–50% longer than fully fed animals. The PHA-4 protein, which was originally described for its role in specifying the pharynx in worm embryos, is a member of the forkhead family of transcription factors, and is very similar to mammalian FOXA proteins[3,4]. In mammals, FOXA proteins have developmental roles, and regulate glucose metabolism later in life[5]. In *C. elegans*, PHA-4 is present from embryo to adult, but its functions in later life were largely unknown.

To examine PHA-4 function in adult worms, Panowski and colleagues (page 550)[1] used genetic manipulations to inactivate the *pha-4* gene in adult worms, without affecting its embryonic activity. Intriguingly, they found that *pha-4* mutants did not respond to dietary manipulations, showing a similar median adult lifespan at all tested food concentrations; however, other processes such as food ingestion and feeding behaviour appeared normal. By inference, therefore, PHA-4 function in adult worms is to respond to dietary changes, and to increase survival under conditions of dietary restriction.

The authors also found that the requirement for PHA-4 was very specific. Systematic knockdown of the other 14 forkhead transcription factors expressed in the worm — including DAF-16/FOXO — had little effect on longevity induced by restricting dietary intake. The DAF-16/FOXO transcription factor works in another well-known longevity pathway — that of insulin/IGF signalling, which dramatically influences lifespan in worms, flies and mice[6]. A modest reduction of insulin/IGF signalling activates DAF-16/FOXO, which mediates a twofold extension of worm lifespan. Similarly, inactivation of DAF-16 completely abolishes insulin/IGF-signalling-mediated longevity[6].

Panowski *et al.* found that, by contrast, mutants lacking DAF-16/FOXO still showed a normal response to dietary restriction, indicating that longevity induced by restricted food intake is DAF-16/FOXO independent. Conversely, PHA-4 had little or no effect on longevity induced by insulin/IGF signalling, because with a reduction in this pathway animals lived longer in the presence or absence of PHA-4 function. Thus, PHA-4/FOXA seems to specifically affect dietary-restriction-mediated longevity, whereas DAF-16/FOXO is involved in regulating longevity induced by insulin/IGF signalling. The authors also found that a conserved nuclear factor called SMK-1 is required for longevity in both pathways[1,7].

Other observations by these authors[1] also support a role for a PHA-4-mediated effect of dietary restriction. First, dietary restriction increased the levels of PHA-4 messenger RNA, indicating a cause-and-effect relationship. In addition, dietary restriction triggered the expression of several genes encoding superoxide-dismutase enzymes, which protect animals from oxidative damage — a cause of ageing. The expression of these genes in response to dietary restriction required PHA-4 activity, suggesting that they could be transcriptional targets of PHA-4.

In adult worms, PHA-4 is found in the intestine, gonad and a handful of neurons. Nevertheless, it influences survival of the whole organism, probably by controlling the production of hormones in some of these tissues, which then signal throughout the body. Indeed, mammalian FOXA regulates glucagon[8], a hormone that is released during fasting in higher organisms. Although worms lack a glucagon homologue, they have several hormone-like proteins that could subsume a similar role.

Like PHA-4, SKN-1 also functions early in embryonic development, where it specifies the formation of the intestine and related tissues[9]. Later in life, it helps to protect animals from oxidative stress[10]. This transcription factor is related to NRF2 transcription factors, which perform similar functions in mammals[11]. Mutant worms lacking SKN-1 have a shortened lifespan[10], but a direct role for SKN-1 in

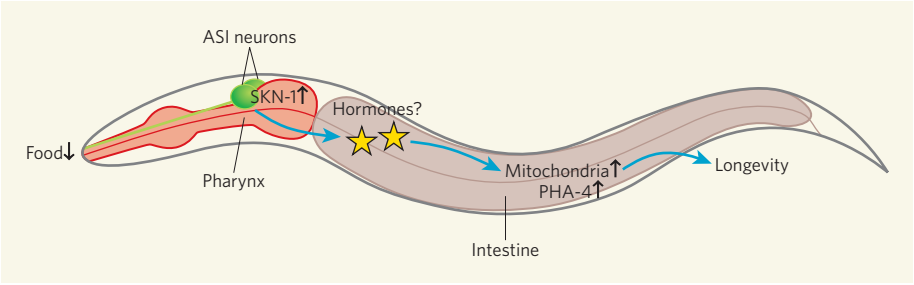


Figure 1 | Dietary restriction and longevity in *C. elegans*. Two studies[1,2] show that, in response to dietary restriction, the activities of SKN-1 and PHA-4 gene-transcription factors increase (black arrows). Neuronal, but not intestinal, SKN-1 mediates longevity in response to reduced dietary intake, where it triggers the release of unidentified hormones (stars) from the pair of ASI neurons to increase mitochondrial activity throughout the body. The PHA-4 transcription factor may also induce hormonal production in the tissues where it is expressed — neurons, intestine and gonad.

mediating dietary restriction was unknown.

Bishop and Guarente (page 545)[2] discovered that a lack of SKN-1 abolishes dietary-restriction-induced longevity over a wide range of food concentrations. Again, the role of SKN-1 in mediating longevity through dietary restriction was specific, because its removal had little effect on lifespan enhanced through a reduction of insulin/IGF signalling. In adult worms, a different form of SKN-1 resides in the intestine from that found in the single pair of neurons known as the ASIs[10]. The authors found that the SKN-1 from the ASI neurons, and not from the intestine, was required for dietary-restriction-induced longevity (Fig. 1), and that restriction of dietary intake promoted increased SKN-1 expression specifically in these two cells. The ASIs are neurosensory cells that integrate cues from the environment and produce various hormones that are relayed to the whole body[12,13]. This indicates that SKN-1-dependent hormonal signals that are released from the ASIs coordinate organism-wide physiological responses to dietary restriction.

Indeed, Bishop and Guarente[2] observed global metabolic changes in response to dietary-restriction-induced activation of *skn-1*, indicative of an organismal response. In particular, restricting dietary intake increased the rate of oxygen consumption — a marker of mitochondrial respiration — in a SKN-1-dependent manner. Mitochondria are the powerhouses of the cell, deriving cellular energy from various metabolites and consuming oxygen in the process. The authors found that, when mitochondrial activity was chemically inhibited, the beneficial effects of dietary restriction on survival were also abolished. Evidently, dietary restriction seems to increase mitochondrial activity, presumably to efficiently extract energy from a limited amount of food. How increased mitochondrial activity might promote longevity is not known. Conceivably, it may trigger mechanisms that protect cells against oxidative stress, or stimulate turnover of damaged cellular components.

Together, these findings[1,2] indicate that dietary restriction activates a highly regulated process, rather than passive metabolic changes. A role for PHA-4 and SKN-1 transcription factors in mediating the effects of dietary restriction on longevity is exciting because, so far, only a handful of molecules have been described that regulate physiological responses to dietary restriction. Moreover, because *pha-4* and *skn-1* genes are evolutionarily conserved, similar mechanisms may hold in higher organisms.

These findings also raise a host of other questions. Although both PHA-4 and SKN-1 are required for dietary-restriction-induced longevity, neither is truly sufficient. Are other factors involved? Must PHA-4 and SKN-1 work together? Moreover, what is the nature of the hormonal pathway that coordinates the effects of dietary restriction? Do the vertebrate counterparts of these transcription factors also have a role in regulating survival? Potentially, the answers to these questions may illuminate the path to increased human health and longevity. ■

Adam Antebi is at the Huffington Center on Aging and the Department of Molecular and Cellular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA.
e-mail: aantebi@bcm.edu

1. Panowski, S. H., Wolff, S., Aguilaniu, H., Durieux, J. & Dillin, A. *Nature* **447,** 550–555 (2007).
2. Bishop, N. A. & Guarente, L. *Nature* **447,** 545–549 (2007).
3. Horner, M. A. *et al. Genes Dev.* **12,** 1947–1952 (1998).
4. Kalb, J. M. *et al. Development* **125,** 2171–2180 (1998).
5. Friedman, J. R. & Kaestner, K. H. *Cell. Mol. Life Sci.* **63,** 2317–2328 (2006).
6. Kenyon, C. *Cell* **120,** 449–460 (2005).
7. Wolff, S. *et al. Cell* **124,** 1039–1053 (2006).
8. Zhang, L., Rubins, N. E., Ahima, R. S., Greenbaum, L. E. & Kaestner, K. H. *Cell Metab.* **2,** 141–148 (2005).
9. Bowerman, B., Eaton, B. A. & Priess, J. R. *Cell* **68,** 1061–1075 (1992).
10. An, J. H. & Blackwell, T. K. *Genes Dev.* **17,** 1882–1893 (2003).
11. Zhang, D. D. *Drug Metab. Rev.* **38,** 769–789 (2006).
12. Bargmann, C. I. & Horvitz, H. R. *Neuron* **7,** 729–742 (1991).
13. Li, C. *Parasitology* **131** (Suppl.), S109–S127 (2005).

HIGH-TEMPERATURE SUPERCONDUCTIVITY

Local pairs and small surfaces

Stephen R. Julian and Michael R. Norman

Mapping out the strange territory of high-temperature superconductors has proved a challenge. In the latest tour de force, two experiments take big steps forward, in complementary directions, to chart the lie of the land.

More than 20 years after they were first discovered, high-temperature superconductors remain fundamentally baffling. In superconductors, conduction without electrical resistance arises through the pairing of electrons so as to overcome obstacles to current flow. But whereas the better-understood 'conventional' superconductors confine their superconducting behaviour to temperatures within a few degrees of absolute zero, certain metallic oxides of copper conduct without electrical resistance at temperatures up to 150 kelvin. Two papers in this issue[1,2] substantially advance our understanding of these bizarre materials. First, in a beautiful study, Doiron-Leyraud *et al.* (page 565)[1] describe the emergence of the classic signature of a metal, a Fermi surface, in a high-temperature superconductor. Second, Gomes *et al.* (page 569)[2] take the most probing look yet at what happens to the electron pairs at temperatures above the transition from the high-temperature superconducting state (Box 1, overleaf).

Doiron-Leyraud and colleagues' measurements[1] have an echo of the old-fashioned physics of metals; however, the small size of the Fermi surface they observe strongly suggests that entirely new physics is in play.

The Fermi surface is named after the Italian physicist Enrico Fermi, who in 1926 suggested that current-carrying electrons fill up available energy states in a metal rather like water fills a lake. The Fermi surface — a strange, often geometrically beautiful, structure that separates filled states from unfilled states in an abstract energy space known as *k*-space — is the shoreline of this lake. For simple metals, its size and shape depend only on the electron density and the crystal structure of the metal.

In the 1960s, the combination of measurements in high magnetic fields and computer-aided calculations allowed the Fermi surfaces of many metals to be mapped out. It became apparent that the existence of the Fermi surface explains more than just the universal features of metals, such as the dependence of their electrical resistance on temperature. Its detailed

NATURE|Vol 447|31 May 2007

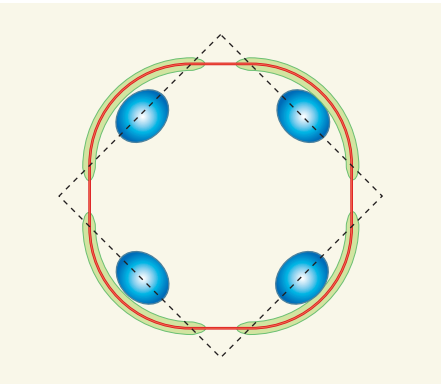


Figure 1 | Search for the Fermi surface. The Fermi surface is depicted in an abstract energy space known as *k*-space (here, a two-dimensional projection of the actual three-dimensional space), and maps out the areas of occupied electron energy states in a metal. The conventional theory of metals demands a large Fermi surface for high-temperature superconductors (red line), as indeed is found in overdoped cuprate samples, which have large numbers of charge carriers. For underdoped samples, the picture is more complex, with photoemission measurements indicating a Fermi surface of truncated 'arcs' (green). Alternatively, smaller 'pockets' (blue) might be centred along special lines in *k*-space (dotted black line). These pockets could emerge from hidden order that introduces an additional periodicity in the material's crystal structure. The enclosed areas of the arcs and the pockets have been adjusted to match the measurements of Doiron-Leyraud *et al.*[1] on their extremely pure yttrium barium copper oxide sample.

shape also accounts very precisely for other, decidedly non-universal, features, such as the dependence of the resistance on the magnetic field. This dependence can be upwards, downwards, oscillatory and even — in the so-called Hall effect — sideways, where a voltage arises at right angles to the direction of the current flow. In particular, the oscillatory Hall signal directly reflects the size of the Fermi surface.

Attempts in the early 1990s to see the Fermi surface in high-temperature superconductors by such means produced controversial sightings. As a consequence, newer techniques such as photoemission spectroscopy, in which light is used to 'knock' out the electrons, and so sketch out the Fermi surface, were drafted into the hunt. As these studies progressed, the supposed Fermi surface seemed to melt away. Gradually, the feeling emerged that perhaps there was no Fermi surface at all in high-temperature superconductors — a truly novel situation for metallic materials.

Doiron-Leyraud *et al.*[1] have dramatically changed the story with 'smoking gun' evidence for a Fermi surface in a high-temperature superconductor: the observation of a Hall resistance that oscillates up and down with increasing magnetic field. Such oscillations are quantum-mechanical in origin, and arise from the circular motion of a metal's electrons in a magnetic field. Observing them requires very pure samples. The sample used by Doiron-Leyraud *et al.* is the result of 19 years' work on one family of cuprate superconductors, yttrium barium copper oxide[3]. This particular sample is special even by the standards of those efforts: it has what is known as an ortho-II structure, in which all the oxygen atoms in the crystal are found in a well-ordered superstructure.

So what can we learn from the appearance, after all this time, of a Fermi surface in a high-temperature superconductor? That depends on its origin, which is a debate likely to dominate the physics of cuprate superconductors in the immediate future. It cannot be the 'large' Fermi surface predicted by the conventional theory of metals: that would produce a much higher frequency of resistance oscillation with increasing magnetic field than is observed. Interestingly, such a large surface has been inferred from photoemission spectroscopy[4] and magneto-transport measurements[5], but only in 'overdoped' cuprates, which have much higher densities of charge carriers and seem to behave more like conventional metals. Owing to its peculiar crystal structure, an underdoped ortho-II sample such as that investigated by Doiron-Leyraud and colleagues[1] wouldn't necessarily have a simple, large Fermi surface. But even the more complicated Fermi surface predicted for the ortho-II cuprate[6] doesn't seem to fit these new experimental observations.

The authors' oscillation measurements tell us the size of the Fermi surface, but not where it is in *k*-space. Because of that, there is a degree of latitude in the physical interpretations of the results. Photoemission results[7] indicate that, in underdoped cuprates, the large Fermi surface found in overdoped samples breaks up into 'arcs' — one of the many novel features of underdoped cuprates that are associated with the presence of a so-called 'pseudogap'. One suggestion made by the authors is that these arcs form small pockets that constitute an enclosed Fermi surface of the observed size (Fig. 1). Alternatively, 'hidden order' associated with the pseudogap might introduce an additional periodicity in the material and cause the large Fermi surface to break up into smaller pockets[8]. A further suggestion is that these small pockets might be a natural consequence of underdoped cuprates also being lightly doped insulators. Finally, a small surface could arise from some unknown physical mechanism in which the magnetic field quenches the long-range order, but local superconductivity persists.

A surprising feature of Doiron-Leyraud and colleagues' measurements[1] is that the oscillatory signal occurred in a Hall resistance that had a negative sign. This is 'wrong' according to conventional ideas: the Hall coefficient is expected to be positive, because carriers are induced in the cuprates by removing electrons. The negative Hall coefficient might be associated with local superconducting vortices that persist in magnetic fields far above that where the resistance becomes finite[9]. But perhaps, as with so many other aspects of cuprate physics, there is something really novel going on that we haven't yet imagined. Certainly, the understanding of how the quantum oscillations of the newly discovered Fermi surface arise will

Box 1 | Breaking up is hard to do

The idea that matter changes phase with rising temperature is familiar: we all know how a solid melts to a liquid when its temperature increases. A similar 'melting' process occurs in superconductors and magnets. One of the many conundrums posed by high-temperature superconductors is how exactly the phase transition out of the superconducting state occurs.

In one important respect, the conventional superconducting and magnetic phases represent opposite ends of a spectrum of phase transitions. When the temperature of a material goes above that at which the macroscopic consequences of magnetism disappear, the atomic moments do not disappear, but are just no longer aligned with one another. The individual moments disappear only at a much higher temperature. In virtually all superconductors, by contrast, the electron pairs thought to be responsible for superconductivity break up at the same critical temperature at which the macroscopic behaviour of zero resistance disappears.

But high-temperature superconductors, it seems, are intermediate between these two limits of phase-transition behaviour. The persistence of pairs in such superconductors above the superconducting transition temperature has been inferred using various probes. But most of these probes are spatially averaged, and how the pairs break up on a local scale as the temperature rises was unknown.

In a novel experiment reported in this issue, Gomes *et al.* (page 569)[2] used a scanning tunnelling probe to follow the pairing process in a cuprate high-temperature superconductor. This proved a challenge, because the thermal vibrations of the probe tip make it difficult for it to remain focused on a single atom. The authors found that the local gap in the superconductor's allowed energy states caused by electron pairing disappeared at a temperature at which the thermal energy was equivalent to a quarter of the size of the local energy gap at absolute zero temperature.

The larger the gap, therefore, the higher the temperature at which the gap persists — in many cases, well above the temperature at which the superconducting behaviour of the bulk material ceases. These measurements clearly imply that at least some of the unusual normal-state properties of the cuprates above their superconducting transition temperature are due to the survival of pairs. Indeed, breaking up is never easy. **S.R.J. & M.R.N.**

be pivotal to solving the mystery of high-temperature superconductivity. ■

Stephen R. Julian is in the Department of Physics, University of Toronto, 60 St George Street, Toronto, Ontario M5S 1A7, Canada. Michael R. Norman is in the Materials Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA.
e-mails: sjulian@physics.utoronto.ca; norman@anl.gov

1. Doiron-Leyraud, N. *et al. Nature* **447,** 565–568 (2007).
2. Gomes, K. K. *et al. Nature* **447,** 569–572 (2007).
3. Liang, R., Bonn, D. A. & Hardy, W. N. *Physica C* **336,** 57–62 (2000).
4. Damescelli, A., Hussain, Z. & Shen, Z.-X. *Rev. Mod. Phys.* **75,** 473–541 (2003).
5. Hussey, N. E. *et al. Nature* **425,** 814–817 (2003).
6. Bascones, E. *et al. Phys. Rev. B* **71,** 012505 (2005).
7. Norman, M. R. *et al. Nature* **392,** 157–160 (1998).
8. Chakravarty, S. *et al. Phys. Rev. B* **63,** 094503 (2001).
9. Xu, Z. A. *et al. Nature* **406,** 486–488 (2000).

EVOLUTIONARY BIOLOGY

Animal personalities

Alison M. Bell

That different people differ in their readiness to take risks is an obvious feature of human personality. Theoretical advances now help in making sense of observations of analogous behaviour in animals.

Personality might seem to require a complexity and subtlety that is unique to humans. But evidence for individual variation in traits that we would recognize as personality, for example aggressiveness in fighting or boldness in the face of a predator, has cropped up in animals ranging from fish to monkeys to squid. Even an individual spider behaves differently from other spiders, through time and in different situations[1]. Wolf *et al.* (page 581 of this issue[2]) now show how such variation in behaviour can make evolutionary sense.

Personality has been difficult to explain from an evolutionary perspective because, at first glance, it could seem maladaptive[3]. An individual that is consistently uninhibited and bold is going to end up eaten by a predator. The optimal animal should be bold only when it makes sense to be bold, and adjust its behaviour when the situation changes. Although animals are legendary for their remarkable 'behavioural plasticity' (think migration or camouflage, for example), there is growing evidence that animals do not always change their behaviour as much as they should. In other words, behavioural plasticity is limited[3].

One possible explanation for this is that individuals should behave consistently if it's simply too hard to undergo a personality transformation. If turning off a general tendency to be aggressive requires time and energy to entirely rewire neural machinery, or to build a physiology that can support a different metabolism, then individuals might be better off sticking to an intermediate strategy[4]. Similarly, if information about the immediate environment is uncertain, then it makes sense just to behave the same way and avoid the risk of making a mistake[5].

This line of reasoning can help to explain why a given individual behaves consistently, but not, for example, why some individuals are always more aggressive than others. Such variation is puzzling, because natural selection will favour individuals with characteristics that perform the best, and less 'fit' individuals will be removed from the population. If a trait is heritable and linked to survival or reproductive success, then evolutionary theory tells us that variation will eventually disappear from the population. But, empirically, we know that personality traits are heritable[6], are linked to fitness[7] and are quite variable.

So how is all this behavioural variation maintained? One way is if the fitness of one strategy depends on the frequency of other strategies in the population[8,9]. Imagine, for example, a group composed entirely of individuals that accumulate resources by guarding them — territorial male birds, for example. An individual using a different strategy — say, dashing in to sneak the resource while a guard is otherwise occupied — would do well in that situation (so long as it is rare), because it would effectively occupy an 'open niche', devoid of competitors.

Alternatively, behavioural variation can be maintained if the best strategy depends on an individual's 'state', which effectively anchors a personality type[8]. This state can be anything from sex or health to body size, and the idea is that an individual should behave consistently so long as its state does not change. This explanation leaves the question of what maintains variation in state.

Wolf *et al.*[2] offer an answer by proposing that an individual's strategy for survival and reproduction — its life-history strategy — is a relatively unchanging state (unlike hunger level, for example), and that individuals adopt different life-history strategies because of fitness trade-offs. Any behaviour that is related to a life-history strategy will be stable through time and differ between individuals with different strategies.

The authors' model starts by assuming that an individual can either reproduce now, but

50 YEARS AGO

"Incorporation of radioactive amino-acids in the proteins of bull spermatozoa" — It is widely held that ribonucleic acid is directly involved in protein synthesis, and there have been several recent demonstrations of the necessity for the presence of ribonucleic acid during synthesis of proteins. In view of this, it seemed to be of interest to examine protein turnover in mature, ejaculated spermatozoa, which apparently contain at most only traces of ribonucleic acid... The absence of the acid from bull semen has been confirmed in the present investigation... It is possible that in this case deoxyribonucleic acid may be involved in the synthesis of proteins... The other possibility would be to regard protein synthesis in spermatozoa as an enzymatic process independent of nucleic acids.

From *Nature* 1 June 1957.

100 YEARS AGO

Mr. Walter Wellman, who proposes to make another attempt to reach the North Pole by means of his airship *America*, has left for Norway, on the way to Spitsbergen, where the balloon will be inflated. In the first week of July there will be trials of the airship until it is demonstrated that it is ready for the voyage... Mr. Wellman has given Reuter's representative the following particulars of his plans:— The airship has been made 18 feet longer and its lifting power increased by 3000 lb., giving a total lifting force of 19,500 lb. The balloon is 184 feet long and 52 feet in its greatest diameter, its cubic volume being 265,000 cubic feet. With the single exception of Count Zeppelin's airship, this is the largest ever built... The total radius of action is believed to be 2500 miles, or double the distance from the base to the Pole and back again. The balloon will not ascend more than 300 feet to 500 feet, and a guide-rope will trail over the surface of the earth.

From *Nature* 30 May 1907.

50 & 100 YEARS AGO

having acquired low-quality resources, or delay reproduction by one year, having acquired high-quality resources. For example, an individual that becomes sexually mature at a young age will have to balance the benefit of early reproduction against the cost of reproducing at a smaller size. Individuals that postpone reproduction must be able to survive to realize their reproductive expectations, and should therefore be generally risk-averse, whereas the opposite is true for those planning to reproduce early. So stable individual differences in risk-taking behaviours can evolve and be maintained when there is a trade-off between early versus late reproduction.

Along the same lines, Stamps[10] argues that another life-history trade-off — this time between growth and mortality — can favour the evolution of personality. Individuals can either opt to grow fast but risk dying young, or instead elect to take the safe route and grow slowly. Because individuals can benefit from growing at a consistent rate, any behaviour or suite of behaviours that is related to growth rate will be consistent through time. And because both the fast-growing and the slow-growing strategy can be maintained in a population when the two strategies have equal fitness owing to a trade-off with mortality[11], behavioural variation can result.

These models[2,10] not only explain why personality occurs, but also predict what form it should take (that is, behaviours that affect the trade-off are likely to co-vary). And they can explain why personality doesn't always evolve[12]. Moreover, an explanation invoking variation in life-history strategies is particularly powerful because life-history trade-offs are something that all living creatures confront — no one can live for ever and produce an infinite number of healthy offspring. Therefore, life-history trade-offs might help to explain why we see something akin to personality in such a wide variety of species.

However, there is room for further theory that allows more dynamic feedback between life-history strategy and behaviour. For example, imagine that a bold and aggressive individual on the 'reproduce early' schedule is successful in acquiring resources. If so, it has effectively changed its state, and its chance of successfully reproducing in the future ('residual reproductive value') has increased. So it is easy to envisage cases in which it would make sense for that individual to switch to the 'delay reproduction' strategy, and change behavioural tactics accordingly. In other words, transitions between life-history strategies might be mediated by personality, and vice versa.

Nonetheless, these models[2,10] offer an adaptive explanation for animal personality and show that its evolution is logically plausible. They should help to convince sceptics that animal personality is not merely an anecdotal topic. Ironically, the search for evolutionary explanations for why individual animals consistently behave differently might also help to answer a fundamental question about ourselves: why do we have unique personalities? ■

Alison M. Bell is in the School of Integrative Biology, University of Illinois, Urbana-Champaign, 505 South Goodwin Avenue, Urbana, Illinois 61801, USA.
e-mail: alisonmb@life.uiuc.edu

1. Gosling, S. D. *Psychol. Bull.* **127,** 45–86 (2001).
2. Wolf, M., van Doorn, G. S., Leimar, O. & Weissing, F. J. *Nature* **447,** 581–584 (2007).
3. Sih, A., Bell, A. M. & Johnson, J. C. *Trends Ecol. Evol.* **19,** 372–378 (2004).
4. DeWitt, T. J., Sih, A. & Wilson, D. S. *Trends Ecol. Evol.* **13,** 77–81 (1998).
5. McElreath, R. & Strimling, P. *Anim. Behav.* **72,** 1135–1139 (2006).
6. van Oers, K., de Jong, G., van Noordwijk, A. J., Kempenaers, B. & Drent, P. J. *Behaviour* **142,** 1185–1206 (2005).
7. Dingemanse, N. J. & Reale, D. *Behaviour* **142,** 1159–1184 (2005).
8. Dall, S. R. X., Houston, A. I. & McNamara, J. M. *Ecol. Lett.* **7,** 734–739 (2004).
9. Wilson, D. S., Clark, A. B., Coleman, K. & Dearstyne, T. *Trends Ecol. Evol.* **11,** 442–446 (1994).
10. Stamps, J. A. *Ecol. Lett.* **10,** 355–363 (2007).
11. Mangel, M. & Stamps, J. *Evol. Ecol. Res.* **3,** 583–593 (2001).
12. Bell, A. M. *J. Evol. Biol.* **18,** 464–473 (2005).

SOLID-STATE PHYSICS

A polariton laser

Leonid V. Butov

Polaritons — particles comprising both light and matter — can form a coherent state, just as light and matter can individually. This fact has now been exploited to make the first room-temperature polariton laser.

At the foundation of modern quantum physics, waves in nature were divided into electromagnetic waves, such as the photon, and matter waves, such as the electron. Both can form a coherent state in which individual waves synchronize and combine. A coherent state of electromagnetic waves is known as a laser; a coherent state of matter waves is termed a Bose–Einstein condensate. But what if a particle is a mixture of an electromagnetic wave and matter? Can such particles form a coherent state? What does it look like? Writing in *Physical Review Letters* Christopoulos *et al.*[1] provide insight into these questions, demonstrating a coherent state of matter–light excitations, or polaritons, in a semiconductor microcavity: the first reported room-temperature polariton laser.

A semiconductor microcavity is built by placing a semiconductor layer between two mirrors. Excitation by an external source, typically a laser, can populate the cavity with different types of particle. The first of these is an exciton, a bound pair of an electron and a 'hole' — an empty state in the band of allowed electron states in the semiconductor. Excitons act as matter particles with a mass equal to the combined mass of an electron and a hole (typically about the vacuum electron mass, m_e), and can form a coherent state similar to a Bose–Einstein condensate.

The second occupant of the cavity is a photon. Although in an empty space a photon is massless, when it is confined in a planar cavity its dispersion relation — the dependence of its energy on its momentum — is modified to the same form as that of a particle with a mass $m = (n\lambda_c/2L)m_e$. Here, n is the refractive index of the cavity medium, L the cavity width and λ_c a constant, the Compton wavelength. For a typical gallium arsenide (GaAs) semiconductor cavity of width 1 micrometre, this mass is about 100,000 times smaller than the electron mass.

If the microcavity is the right width, the energies of the cavity photon and the exciton can be made to match up. When this happens the two mix, forming a new particle. This is a combination of matter and electromagnetic waves — an exciton-polariton, or simply 'polariton'. These polaritons inherit some of the lightness of the cavity photons, and have masses much smaller than m_e.

By varying the parameters of the semiconductor microcavity, several types of coherent state can be obtained. The first forms when the densities of electrons and holes in the cavity medium are so high that the polaritons are essentially destroyed (typically because of dephasing, screening and phase-space filling owing to the carrier Coulomb interaction[2]). In this regime, which is known as weak coupling, the protagonists are the cavity photons and the electron–hole pairs. The system is analogous to a conventional laser: the photon is amplified by stimulating the electron and hole to recombine, emitting a second, coherent photon of the same energy. It is, in fact, a close relative of a type of laser known as a VCSEL (vertical-cavity surface-emitting laser). In contrast to conventional edge-emitting semiconductor lasers, VCSELs emit their beam perpendicular to the active region of the laser, so they can be built in a dense two-dimensional array. Moreover, the beam of a VCSEL is characterized by a lower divergence angle compared with that of the edge-emitting lasers; as a result they are favoured for use in fibre-optic communications.

After the demonstration of lasing in the weak-coupling regime, research moved on to creating a coherent state in the strong-coupling, or polariton, regime. This required prevention of the polariton's destruction before a coherent

COLOUR VISION

Seeing red

André A. Fernandez and Molly R. Morris have revisited a question that evolutionary biologists just can't let go of. Many species of primate have trichromatic colour vision. But what factors drove its evolution?

The authors have tested the 'pre-existing-bias hypothesis' (*Am. Nat.* doi:10.1086/518566; 2007). This holds that the evolution of a trait such as colour vision might have evolved to meet one particular need, but was subsequently favoured by selection for another.

A long-standing idea is that trichromatic vision is of great benefit in foraging — in allowing red or orange food items to be seen more easily in green foliage. Another is that its genesis lies in promoting sexual selection — that intense colour, in particular red as skin or fur coloration, evolved for example as a signal of sexual state or prowess. Fernandez and Morris set out to provide a statistical test of the possibility that selection of this latter putative function was due to a pre-existing bias.

Their approach was to mine the literature to produce 'ancestral state reconstructions' of the incidence among primate groups of colour vision, fur colour, skin colour and form of mating system. The thinking behind the inclusion of this last characteristic was that selection for colourful signals is likely to be greater in gregarious species. These reconstructions were then subjected to tests of whether red fur and skin were more likely to have evolved in the presence of colour vision and gregarious mating systems.

The conclusion that emerges is that trichromatism preceded a tendency towards red coloration and gregarious mating systems — that is, that there was indeed a pre-existing bias towards red (probably, think the authors, for more efficient foraging) that then prompted the coevolution of red skin and fur through sexual selection.

There are obviously plenty of exceptions to the correlation between trichromatism and red coloration. Explanations for such exceptions include lack of suitable genetic variation in the species concerned, lack of the appropriate diet or predator pressure.

Fernandez and Morris add that their research may tie in with work on the evolutionary history of the vomeronasal organ. This is a sensor for scent signalling, and there is evidence that its functional loss in Old World monkeys coincided with the advent of colour vision — prompting the idea that chemical signalling gave way to visual cues.

I. ARNDT/NATUREPL.COM

Tim Lincoln

state could form. This was achieved by using resonant pumping of the cavity with photons at a specific angle[3]. The polaritons created by this excitation effectively scatter to the lowest-energy polariton mode, and can be amplified there through polariton–polariton stimulated scattering. The result was another type of laser, dubbed the polariton parametric amplifier.

The first demonstration of this laser was at cryogenic temperatures. The upper temperature for the polaritons is ultimately limited by the exciton binding energy: a polariton's constituent exciton will dissociate at high temperatures T when k_BT becomes comparable to its binding energy (k_B is Boltzmann's constant). With materials of a suitably high exciton binding energy, however, success was possible at temperatures as high as 220 K (–53 °C)[4].

Christopoulos *et al.*[1] focused their research on creating a coherent polariton state in a semiconductor microcavity using non-resonant pumping in the strong coupling regime. Such a coherent state was proposed[5] and has been realized[6,7] in both GaAs and cadmium telluride (CdTe) microcavities at cryogenic temperatures. Success at room temperature, however, required a new solution and new materials with a high exciton binding energy and exciton–photon interaction strength[4,8]. Gallium nitride (GaN) was Christopoulos and colleagues' material of choice. They achieved the transition to the coherent polariton state at a remarkably low threshold power, an order of magnitude smaller than in the best GaN-based VCSELs.

So is this polariton coherent state similar to a laser, to a Bose–Einstein condensate, or somehow to both? In other words, is it the electromagnetic-wave or the matter-wave component of the polariton that is responsible for establishing the coherent state? This is an issue of more than semantic significance. The fundamentally different nature of a laser and a Bose–Einstein condensate is revealed in, for example, the different critical temperatures below which lasing or condensation occurs. Results so far show that the coherent states of microcavity polaritons disappear when the photon component vanishes in a system of bare excitons, but remain strong when the exciton component vanishes in the weak-coupling regime. It seems that the coherence of the microcavity polaritons arises from the coherence of an electromagnetic field — just as in a laser.

Christopoulos and colleagues' breakthrough opens up intriguing avenues for investigation. First, it provides a model system for fundamental studies of light–matter interaction at elevated temperatures. Second, there is the promise of practical applications of the polariton lasers, for instance in fibre-optic communications. ■

Leonid V. Butov is in the Department of Physics, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0319, USA.
e-mail: lvbutov@physics.ucsd.edu

1. Christopoulos, S. *et al. Phys. Rev. Lett.* **98,** 126405 (2007).
2. Khitrova, G. *et al. Rev. Mod. Phys.* **71,** 1591–1639 (1999).
3. Savvidis, P. G. *et al. Phys. Rev. Lett.* **84,** 1547–1550 (2000).
4. Saba, M. *et al. Nature* **414,** 731–735 (2001).
5. Imamoglu, A. & Ram, R. J. *Phys. Lett. A* **214,** 193–198 (1996).
6. Deng, H. *et al. Science* **298,** 199–202 (2002).
7. Kasprzak, J. *et al. Nature* **443,** 409–414 (2006).
8. Lidzey, D. G. *et al. Nature* **395,** 53–55 (1998).

PROTEIN SCIENCE

Discriminating taste of prions

Witold K. Surewicz

Prions are infectious proteins that are involved in brain-wasting disorders such as mad cow disease. In yeast, specific sequences of amino acids in prions seem to mediate prion propagation and cross-species transmissibility.

Few diseases have generated as much interest and controversy in recent years as transmissible spongiform encephalopathies (TSEs) — a group of fatal neurodegenerative disorders that includes Creutzfeldt–Jakob disease in humans and 'mad cow' disease[1,2]. The protein-only hypothesis of TSE propagation postulates that the infectious pathogen, called a prion, is devoid of the nucleic acids that comprise genes, and is an abnormally shaped version of the normal prion protein, a naturally occurring molecule of unknown function. The infectious prion replicates by binding to its host's normal prion protein and forcing it to take on the abnormal conformation[1]. Once heretical, this notion that proteins alone can be infectious is now rapidly gaining acceptance. Paradoxically, the most compelling evidence for this idea has come not from studying TSEs, but from prion-like entities in yeast and other fungi[3–6].

In this issue (page 556), Tessier and Lindquist[7] present intriguing findings that clarify many of the mechanisms underlying the conversion of yeast prion proteins to an infectious form.

Whereas mammalian prions cause a dreaded brain-wasting disease, infectious prions in yeast change the behaviour (or phenotype) of these simple organisms, with the altered characteristics passing from one generation to the next[3,4]. Thus, yeast prions act as protein-only 'genes'. The most extensively studied yeast prion protein is called Sup35, which, when conformationally modified into an infectious prion, leads to a yeast phenotype that is defective in terminating protein synthesis. The conversion of Sup35, and other yeast prions, to the infectious state is associated with their polymerization into amyloid fibrils — highly organized, thread-like aggregates that are also linked to Alzheimer's disease and other 'protein-misfolding' disorders[3,4].

To investigate the precise mechanisms underlying prion conversion, Tessier and Lindquist[7] synthesized a large library of overlapping peptides derived from the amino-acid sequences of two regions of Sup35 — its N-terminal (N) and its middle (M) regions, jointly denoted ScNM — from the budding yeast *Saccharomyces cerevisiae*. They then used an array assay to analyse the ability of these sequences to bind to and modify the full-length protein (Fig. 1). They found that only a tiny subset of ScNM peptides could capture the ScNM protein from solution, and that these peptides selectively promoted the assembly of amyloid fibrils. The authors conclude that small elements of the Sup35 amino-acid sequence can recognize soluble conformations of the protein with unusual specificity, and that this is sufficient to drive the conversion of soluble proteins to the abnormal prion state.

How can the exquisite sequence specificity observed by Tessier and Lindquist be reconciled with a previous study[8] showing that scrambling the amino-acid sequence in the N-terminal domain of Sup35, which is crucial for protein assembly into amyloid fibrils, does not prevent the formation of heritable prions? Clearly, this previous finding would argue that amino-acid composition, not sequence, is important for prion propagation. Although there is no obvious answer to this conundrum, it is possible that scrambling causes new recognition elements to emerge. This is a plausible scenario, given that almost half of the amino-acid residues in the Sup35 N-terminal domain are glutamine and asparagine. That is, no matter how much the sequence is randomized, there will always be regions rich in these amino acids, which will favour amyloid fibril formation.

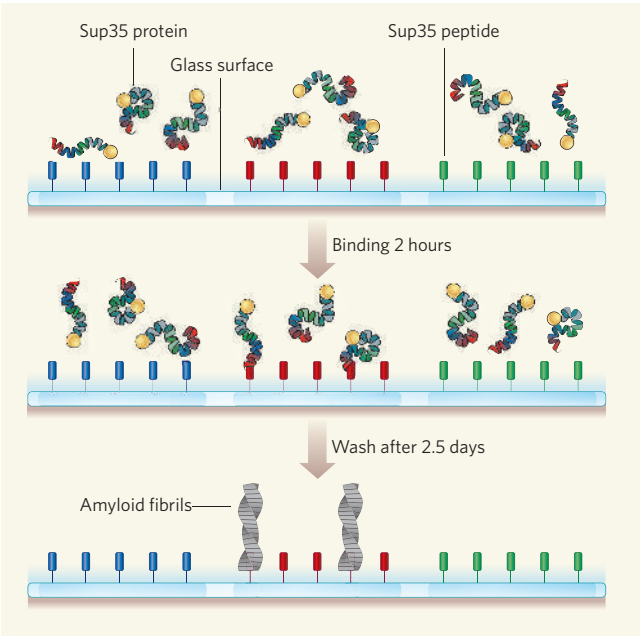


Figure 1 | Principles of the immobilized-peptide array assay. To determine which part of the Sup35 yeast protein is responsible for its susceptibility to transform into an infectious prion, Tessier and Lindquist[7] performed an array assay. They generated peptides of 20 amino-acid residues corresponding to overlapping fragments of Sup35, attached them to a glass surface by covalent immobilization through a linker molecule, and incubated them with a solution containing intact Sup35 proteins. Only a very small subset of the peptides could capture the protein from solution, indicating highly selective recognition (middle panel). The selective-recognition peptides induced the soluble proteins to change into a self-perpetuating amyloid state (bottom panel). Different colours represent individual elements of amino-acid sequence. Yellow balls represent fluorescent labels used for detection.

The most challenging questions in prion research relate to the mechanisms that control the transmission barriers between species. This problem has major public-health implications, because it seems that prions responsible for mad cow disease have crossed over to humans, causing a new variant of Creutzfeldt–Jakob disease[2]. In their study, Tessier and Lindquist have also investigated the issue of prion transmissibility barriers, and their findings are fascinating.

The authors repeated the array experiments using Sup35 sequences (CaNM) from another yeast species, *Candida albicans*. They found that soluble Sup35 proteins from this species bind with great selectivity to their own immobilized peptide sequences, but do not bind to ScNM peptides, and vice versa. A 'species barrier' as observed above, however, can occasionally be crossed if many abnormal prion forms, or strains, of the same protein are present that differ in fibril conformation[4,9]. But how does this happen?

To answer this question, the authors used an engineered chimaeric protein consisting of segments from both ScNM and CaNM (Sc/Ca NM). It is known that, at different temperatures, Sc/Ca NM assembles into fibrils with distinct 'seeding' (amyloid templating) specificities[10]. Tessier and Lindquist[7] found that at 4 °C — the optimum temperature for producing fibrils that selectively seed the amyloid formation of ScNM — the Sc/Ca NM chimaera bound exclusively to immobilized ScNM peptides. At 37 °C, a favourable temperature for the formation of fibrils with CaNM-seeding specificity, only CaNM peptides were recognized by the chimaeric protein. Collectively, these results indicate that small elements in the primary sequence of prion proteins act as extremely selective recognition sites, and play a central role in inducing a switch to the amyloid state. These same sequence elements determine the species-specific seeding activities, and govern the formation of distinct prion strains.

Although these findings provide a conceptual basis to explain some of the puzzles of prion biology, many questions remain. For example, is the temperature-dependent, species-specific recognition of the chimaeric protein due to the effect of temperature on the conformation of the chimaera monomer, that of the early oligomeric species present in solution, or that of the immobilized peptides? Are the authors' findings applicable to all prion proteins? This is an especially pressing question because the Sc/Ca NM chimaera is an artificial entity, and so may act differently from naturally occurring prions. Furthermore, unlike largely unstructured yeast prions, mammalian prion proteins contain a folded domain that might obstruct the exposure of potential recognition sites. Despite these questions, the methodology used by Tessier and Lindquist[7], as well as their impressive findings, will open up avenues in prion research and stimulate further investigation of other prions and non-infectious amyloids. ■

Witold K. Surewicz is in the Department of Physiology and Biophysics, Case Western Reserve University, Cleveland, Ohio 44106, USA.
e-mail: witold.surewicz@case.edu

1. Prusiner, S. B. *Proc. Natl Acad. Sci. USA* **95,** 13363–13383 (1998).
2. Collinge, J. *Annu. Rev. Neurosci.* **24,** 519–550 (2001).
3. Wickner, R. B. *et al. Genes Dev.* **18,** 470–485 (2004).
4. Chien, P., Weissman, J. S. & DePace, A. H. *Annu. Rev. Biochem.* **73,** 617–656 (2004).
5. King, C.-Y. & Diaz-Avalos, R. *Nature* **428,** 319–323 (2004).
6. Tanaka, M., Chien, P., Naber, N., Cooke, R. & Weissman, J. S. *Nature* **428,** 323–328 (2004).
7. Tessier, P. M. & Lindquist, S. *Nature* **447,** 556–561 (2007).
8. Ross, E. D., Edskes, H. K., Terry, M. J. & Wickner, R. B. *Proc. Natl Acad. Sci. USA* **102,** 12825–12830 (2005).
9. Jones, E. M. & Surewicz, W. K. *Cell* **121,** 63–72 (2005).
10. Chien, P., DePace, A. H., Collins, S. R. & Weissman, J. S. *Nature* **424,** 948–951 (2003).

OBITUARY

Frank H. Westheimer (1912–2007)

Pioneering chemist at the interface with biology.

S. MITCHELL/HARVARD UNIV. NEWS OFFICE

Today, we recognize that the complexities of biology can be explained by understanding the chemical and physical properties of the molecules, both large and small, that interact to mediate life's processes. However, we should never forget the intellectual foundation that allows us to approach difficult biological problems in this way. Frank Westheimer, who died on 14 April at the age of 95, demonstrated that chemists have a unique advantage in deciphering the key processes of biology. Westheimer used his insight to describe the workings of enzyme-catalysed reactions, providing essential blueprints for future investigations of the biochemical foundations of life.

Westheimer was born in Baltimore, Maryland, on 15 January 1912. He graduated from Dartmouth College in 1932 and received his PhD in chemistry from Harvard University in 1935. Following a postdoctoral year at Columbia University, he joined the chemistry department at the University of Chicago, where he worked for nearly 20 years. However, this period was interrupted by the Second World War, when he was a supervisor at the National Explosives Research Laboratory. In 1953, he returned to Harvard, where he remained until his retirement in 1983. He received the US National Medal of Science in 1986.

Westheimer began his career as a physical organic chemist, and one of his most notable achievements in this area was the invention of molecular mechanics in the late 1940s. This is a mathematical approach in which the bending and stretching parameters of molecules are used to predict the rates of chemical reactions. The manual calculation efforts required were considerable, but molecular mechanics (and its sibling, quantum mechanics) is now used routinely as a computational method to predict the course of enzyme-catalysed reactions.

While at Chicago, Westheimer started working in biochemistry; he was the first to apply the rigorous analysis of physical organic chemistry to enzymes. Because he chose not to focus on a single problem, it is difficult to identify an area of biochemistry that Westheimer has not influenced. His first breakthrough was in the early 1950s, when he used isotopic labelling to study the oxidation of alcohol (ethanol) by the enzyme alcohol dehydrogenase. He discovered that a direct transfer of hydrogen occurs between the substrate and the enzyme cofactor — a then surprising result that demonstrated the power of his chemical approach. Today, this study could be simply performed using nuclear magnetic resonance (NMR) spectroscopy. But NMR wasn't available at the time, so Westheimer's analyses were accomplished by 'brute force': the organic molecules were combusted to produce water, which was converted to hydrogen gas. The isotopic content of that gas was then determined by mass spectrometry.

Westheimer also demonstrated that this enzymatic hydrogen transfer is stereospecific — the active site distinguishes between two chemically equivalent hydrogens in ethanol, transferring just one of them to the cofactor. Nowadays, it is understood that chemically equivalent hydrogen atoms in molecules can be differentiated in the chiral environment of an enzyme active site. But Westheimer's study was the first demonstration of this effect, and his results were unexpected.

Shortly after moving to Harvard, Westheimer went on to investigate another important class of biochemical transformation: phosphoryl transfer reactions. These are essential for many biological processes, including the production and use of metabolic energy, the storage of genetic information, and signal transduction. The central chemical event in these processes is the cleavage of bonds in phosphate ester molecules, which come in two varieties: monoesters (as found in the energy-storage molecule ATP) and diesters (such as those found in DNA). Westheimer established that the mechanisms of cleavage in phosphate diesters and monoesters are not the same, so providing a crucial understanding of how enzymes catalyse these diverse reactions.

With typical eclecticism, Westheimer made another ground-breaking discovery in the 1960s by determining how enzyme active sites stabilize carbanions — unstable organic anions that often occur as intermediates in enzyme-catalysed processes. Focusing on the reaction mediated by the acetoacetate decarboxylase enzyme, he showed that the carbanion intermediate is stabilized by a positively charged group (a β-iminium ion) that forms when the substrate reacts with a lysine amino acid side chain in the active site. He also established the importance of electrostatic effects in modulating the reactivities of chemical groups in active sites. In the case of acetoacetate decarboxylase, the active-site environment decreases the basicity of the essential lysine residue, a necessary effect if the enzyme is to work at neutral pH.

Westheimer continued to be intrigued by the cleavage of phosphate esters, and I joined his laboratory as a graduate student in the early 1970s as part of his research effort in that area. One day, I realized that I had inadvertently crystallized a crucial enzyme, so I asked Frank to come with me from his office to the microscope in the laboratory. Despite protestations that he was busy, he followed me, and I will never forget his smile when I told him what he was looking at. To a chemist, crystals are both measures of purity and objects of beauty. To Frank, crystals of a molecule as complex as an enzyme were a special joy.

Westheimer's achievements extended beyond academia. In 1965, he chaired the US National Academy of Sciences Committee for the Survey of Chemistry. This body issued the influential Westheimer Report, which recommended increases in US federal funding for the chemical sciences and highlighted the importance of the emerging field of biochemistry. Westheimer also served as a science adviser to US President Lyndon Johnson from 1967 to 1970.

I last saw Frank three years ago at his home in Cambridge, Massachusetts. Even at 92, he was intensely interested in my science, just as he had been when I was a graduate student. My plan was to give a private seminar about my most recent work, but I didn't get far before he was reminiscing about his career and how science had changed. As he observed in my work, X-ray crystallography is now used routinely to obtain enzyme structures, providing once unimaginable information for determining their modes of action. It was humbling to recall that his pioneering work on alcohol dehydrogenase was achieved without even the benefit of NMR. But although experimental tools have changed, the intellectual guidelines established by Westheimer for understanding the chemical basis of biology will endure for ever.

John A. Gerlt
John A. Gerlt is in the Departments of Biochemistry and Chemistry, University of Illinois, 600 South Mathews Avenue, Urbana, Illinois 61801, USA.
e-mail: j-gerlt@uiuc.edu

ECOLOGY

Global warming and amphibian losses

Arising from: J. A. Pounds *et al.* *Nature* **439,** 161–167 (2006)

Is global warming contributing to amphibian declines and extinctions by promoting outbreaks of the chytrid fungus *Batrachochytrium dendrobatidis*? Analysing patterns from the American tropics, Pounds *et al.*[1] envisage a process in which a single warm year triggers die-offs in a particular area (for instance, 1987 in the case of Monteverde, Costa Rica). However, we show here that populations of two frog species in the Australian tropics experienced increasing developmental instability, which is evidence of stress[2–4], at least two years before they showed chytrid-related declines. Because the working model of Pounds *et al.* is incomplete, their test of the climate-linked epidemic hypothesis could be inconclusive.

Biotic or abiotic agents may cause stress. During amphibian growth and development, disease or physical stressors can increase the asymmetry of limb lengths[3,4]. Pounds *et al.* propose a mechanism for chytrid outbreaks, in which global warming favours the pathogen directly by shifting microscale temperatures towards its optimum. Unusual climatic conditions, however, may also stress amphibians[5], potentially altering development and raising susceptibility to chytridiomycosis, as well as other diseases[6]. During an epidemic in New South Wales, Australia, in 1999, the chytrid was found in only seven of 32 moribund frogs (*Limnodynastes tasmaniensis*)[7].

To investigate stress, we analysed limb asymmetries in populations near Paluma, in northern Queensland, Australia, just before their decline, and in non-declining 'control' populations of the same species. We measured limb elements of 164 adult males, including 112 *Litoria nannotis* (38 'control' and 74 'pre-decline') and 52 *L. genimaculata* (23 'control' and 29 'pre-decline'). The frogs were obtained during 30 samplings at different times (12 'control', Fig. 1a; and 18 'pre-decline', Fig. 1b).

At Paluma between April and September 1991, *L. nannotis* collapsed and disappeared; *L. genimaculata* declined within 3 months afterwards but later recovered[8]. The controls, from 12 tropical wet forest sites, were obtained during nine years that preceded the die-offs at Paluma by an average of 16 years.

Using a blind protocol[9], we measured every frog three times. We then examined asymmetry in relation to body size, and, using analysis of variance, partitioned the variation representing the different kinds of asymmetry and measurement error[2,9]. To compare samples, we examined the total unsigned asymmetry per individual (the difference between the left and right forelimbs, plus the corresponding value for the hindlimbs).

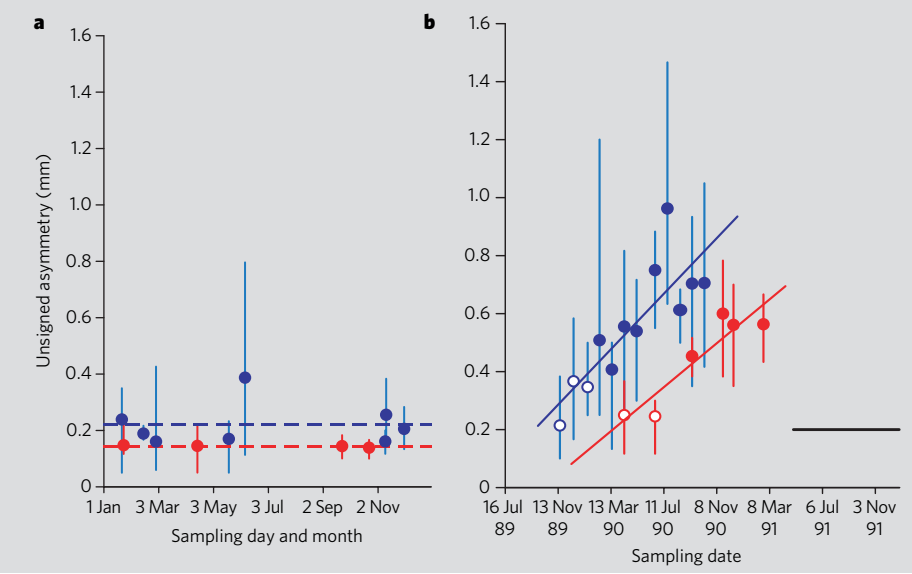


Figure 1 | Limb asymmetry in frog populations that soon declined, compared with non-declining 'control' populations. **a**, Asymmetry does not vary significantly among control samples of *Litoria nannotis* (blue) or *L. genimaculata* (red) (Kruskal-Wallis tests; $P > 0.165$). Sampling years are pooled. Circles are sample means; error bars are ranges of individual values ($3 \le n \le 9$). Dashed lines are pooled means. **b**, Asymmetry increases over time in pre-decline samples (symbols as in **a**). Filled circles are means that are significantly higher than the corresponding pooled controls (Mann–Whitney U-tests, Bonferroni-corrected for multiple comparisons; $P < 0.05$). Open circles are means that are not significantly higher than controls. Horizontal black bar indicates the time period of the local declines.

We found that asymmetry was greater in the pre-decline samples than in the controls (Fig. 1). It was not correlated with body size, which varied little. We found no directional asymmetry, and measurement error was small relative to unsigned asymmetry. At Paluma, the unsigned asymmetry increased over time, becoming significantly higher relative to the controls at 15–19 months before the declines. The frogs reach adult size in 1–2 years, so stress probably began to increase at least two years before these die-offs.

Irrespective of whether this increasing stress reflects disease or changes in the envi-

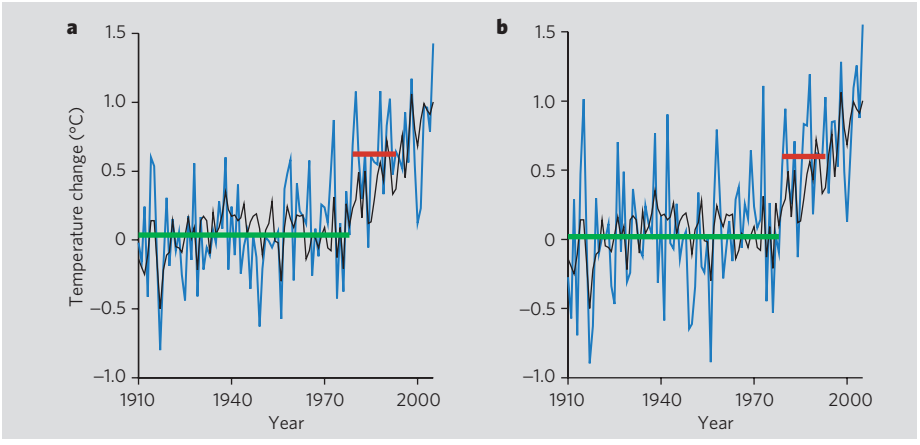


Figure 2 | Temperature changes and the period of major frog die-offs in Queensland, Australia. Annual mean surface temperature for global land areas (black line) is correlated with the corresponding mean (blue line) for **a**, Australia, and **b**, Queensland ($r \ge 0.71$, $P < 0.0001$). Values are departures from a baseline mean for 1910–39. For both regions, the mean for the period of die-offs (red bar) was about 0.6 °C above the baseline, whereas the mean for previous years (green bar) was not significantly increased. During 1988–91, just before the declines at Paluma, the means for these regions were 0.82 °C for Australia and 0.65 °C for Queensland. Data are from the Australian Bureau of Meteorology; details are available from the authors.

ronment, the chytrid was probably not the only agent involved. According to the extinction-wave hypothesis, which has long been controversial[10–12], this pathogen spreads gradually across large regions, causing population crashes within 4–6 months of invading a new site[13,14]. The association of die-offs with previous stress, however, runs counter to this model. It is possible that shifting conditions facilitated the chytrid's invasion, its irruption or its emergence from a saprophytic, quiescent or non-lethal parasitic state.

The data from Queensland cover too few independent die-offs to test for an association with climate change[10], but the period of the major declines[13,15] coincides with dramatic regional warming, which, on diverse spatial scales, mirrors the global trend (Fig. 2). It has been claimed[15] that shifts in temperature or precipitation can be ruled out as factors in these losses, but this claim is undermined by the choice of variables, stations and seasonal timeframes used in that particular study. The patterns from the Australian tropics agree with the climate-linked epidemic hypothesis[1,5,6,11].

Nevertheless, we question the analysis of Pounds *et al.*, which so far provides the only geographically broad test of this idea. Contrary to their working model[1], our Figs 1, 2 indicate that multiyear warm periods may be more important in amphibian declines than single warm years. By focusing on the latter, the authors' test could be inconclusive.

Ross A. Alford*, Kay S. Bradfield*†, Stephen J. Richards*‡

*School of Marine and Tropical Biology, James Cook University, Townsville, Queensland 4811, Australia

e-mail: ross.alford@jcu.edu.au

†Present address: Durell Wildlife Conservation Trust, Les Augres Manor, Trinity, Jersey JE3 5BP, UK

‡Present address: South Australian Museum, North Terrace, Adelaide, South Australia 5000, Australia

Received 13 February; accepted 10 May 2007.

1. Pounds, J. A. *et al. Nature* **439,** 161–167 (2006).
2. van Dongen, S. *J. Evol. Biol.* **19,** 1727–1743 (2006).
3. Parris, M. J. & Cornelius, T. O. *Ecology* **85,** 3385–3395 (2004).
4. Söderman, F., van Dongen, S., Pakkasmaa, S. & Merilä, J. *Oecologia* **151,** 593–604 (2007).
5. Reading, C. J. *Oecologia* **151,** 125–131 (2007).
6. Kiesecker, J. M., Blaustein, A. R. & Belden, L. K. *Nature* **410,** 681–684 (2001).
7. Burgin, S., Schell, C. B. & Briggs, C. *Acta Zool. Sinica* **51,** 344–348 (2005).
8. Richards, S. J. & Alford, R. A. *Aust. J. Zool.* **53,** 229–236 (2005).
9. Alford, R. A., Bradfield, K. & Richards, S. in *Declines and Disappearances of Australian Frogs* (ed. Campbell, A.) 34–43 (Environment Australia, Canberra, 1999).
10. Alford, R. A. & Richards, S. J. *Conserv. Biol.* **11,** 1026–1029 (1997).
11. Bosch, J., Carrascal, L. M., Durán, L., Walker, S. & Fisher, M. C. *Proc. R. Soc. B* **274,** 253–260 (2007).
12. Pounds, J. A., Carnaval, A. C., Puschendorf, R., Haddad, C. F. & Masters, K. L. *Science* **314,** 1541–1542 (2006).
13. Laurance, W. F., McDonald, K. R. & Speare, R. *Conserv. Biol.* **10,** 406–413 (1996).
14. Lips, K. R. *et al. Proc. Natl Acad. Sci. USA* **103,** 3165–3170 (2006).
15. Laurance, W. F. *Biol. Conserv.* **77,** 203–212 (1996).

Competing financial interests: declared none.
doi: 10.1038/nature05940

ECOLOGY

The proximate cause of frog declines?

Arising from: J. A. Pounds *et al. Nature* 439, 161–167 (2006)

Pounds *et al.*[1] argue that global warming contributes to amphibian declines by encouraging outbreaks of the chytrid fungus *Batrachochytrium dendrobatidis*. Although our findings agree with the climate-linked epidemic hypothesis[1–4], this pathogen is probably not the only proximate factor in such cases: in the Trasimeno Lake area of Umbria in central Italy, for example, the water frog *Rana lessonae* first declined in the late 1990s, yet chytridiomycosis was not observed until 2003 (refs 5, 6). Here we show that the chytrid was common there throughout 1999–2002, in a previously unknown form that did not cause disease. We therefore think that the focus by Pounds *et al.* on a single pathogen is hard to justify because the host–parasite ecology is at present so poorly understood.

Chytridiomycosis is thought to follow a simple progression, in which the zoospores encounter amphibian skin and quickly give rise to sporangia, which produce new zoospores[7]. The disease progresses as these reinfect the host. Neither resting spores nor saprobic forms of the chytrid have been reported, and even comparatively resistant hosts, once infected, are expected to harbour sporangia and zoospores. Evidence that the fungus can persist without these stages would suggest that the life cycle is not fully understood and that certain conditions may trigger a switch in modes.

During the European heatwave of 2003 (ref. 8), the water level in Trasimeno Lake fell by 183 cm and *R. lessonae* again declined, whereas the more stress-resistant *R. esculenta* remained comparatively stable[6]. (The latter is a hybrid of *R. lessonae* and *R. ridibunda* in this hybridogenetic complex[5,6].) Meanwhile, a parasitic mesomycetozoean protist that is prevalent in *R. lessonae* (*Amphibiocystidium ranae*)[5,9] spread over a larger area, and we observed chytridiomycosis for the first time — in 18% of the *R. lessonae* sampled[6].

The chytrid, however, was present long before this. During 1999–2002, we examined 400 water frogs in detail, using stereomicroscopy of epidermal surfaces, followed by light and electron microscopy of skin in histological preparations. We found no signs of chytridiomycosis: no sporangia, zoospores, or clinical symptoms[6]. Every year, though, we observed spherical, unicellular organisms, often with a thick wall, in 20–40% of the frogs sampled (Fig. 1). Usually attached to the skin surface, these microbes were confined to minute patches (80–120 μm across) that were easily missed in routine biopsies but appeared in stereomicroscopy as light-coloured pinpoints. Using the same diagnostic tests as before[6], we identified the organisms as a form of *B. dendrobatidis* (see Methods). Their size and other characteristics are suggestive of encysted zoospores: they may embody a resting spore, a saprobe or a parasitic form that is conditionally non-pathogenic.

These findings have several implications. They support the idea that climate change, perhaps combined with other aspects of environmental deterioration such as chemical pollution[10], contributes to amphibian

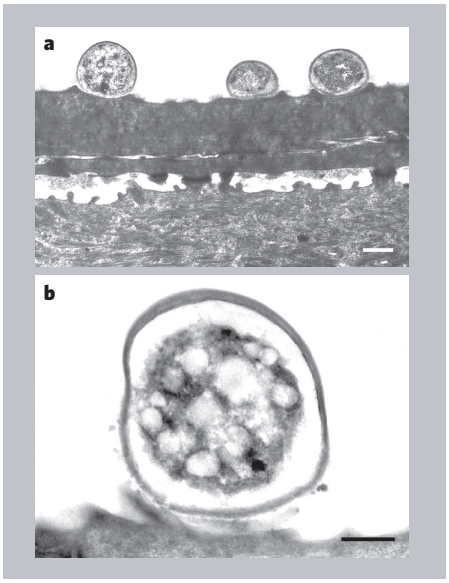


Figure 1 | A newly discovered form of the chytrid fungus *Batrachochytrium dendrobatidis* on the skin of the water frog *Rana lessonae*. a, Transmission electron micrograph of a cross-section of skin showing three of these unicellular organisms adhering to the epidermal surface (scale bar, 1 μm). Although usually attached in this way, they were occasionally intercalated in the superficial epidermis. **b,** Higher-magnification micrograph (scale bar, 0.5 μm) showing a thick cell wall. Sporangia, zoospores and other signs of chytridiomycosis were absent from our samples throughout the period 1999–2002, when the new form was common, but were readily observable in 2003.

declines[1–4,11]. They run counter to the idea[12] that the chytrid simply invades a locality and causes immediate die-offs. Just as this fungus can inhabit areas without causing declines[3,11,13], it may be present without causing disease. Moreover, characteristics of the new form suggest that it could persist in the environment long-term. If so, studying the chytrid largely on amphibian hosts might sketch an inaccurate picture of its spatiotemporal dynamics. This pathogen seems to have emerged recently, but the claim[12] that it is spreading gradually across large geographic regions is arguable[3].

It is also unclear whether climate-related amphibian declines are mediated mainly through outbreaks of the chytrid. This fungus may cause important mortality[12,14] but is not always implicated in such losses[2,11]. Nor is it the only parasite associated with die-offs[2,15]. The white tail tips and other symptoms of tadpoles (*Ptychohyla hypomykter*) that died by the thousands in Guatemala in 2002 (ref. 15) suggest an unidentified pathogen. Although some infections may be secondary, we found that the prevalence of *A. ranae* in 1999–2000, when the chytrid was present but chytridiomycosis was absent, was 49% in the declining *R. lessonae*, compared with 16% in the non-declining *R. esculenta*[5].

By causing stress to the animals, environmental change may raise their susceptibility to various pathogens, many of which are relatively unknown. The exclusive focus by Pounds *et al.* on the chytrid is therefore counterproductive to developing a broader understanding of host–parasite ecology and its role in amphibian losses.

Methods. The molecular analyses and immunohistochemical assays used to identify *B. dendrobatidis* in water frogs have been described[6]. To identify the new form, we examined sections of frog skin taken from the light-coloured pinpoints described in the text. For DNA sequencing, genomic DNA was extracted from these sections and amplified by PCR; for the indirect immunoperoxidase test, we incubated the sections with a polyclonal antibody against *B. dendrobatidis* (provided by A. D. Hyatt) that specifically labelled these unicellular organisms. Both tests demonstrated the presence of the chytrid.

Ines Di Rosa, Francesca Simoncelli, Anna Fagotti, Rita Pascolini
Dipartimento di Biologia Cellulare e Ambientale, Università di Perugia, Via Pascoli 1, Perugia 06123, Italy
e-mail: inesdir@unipg.it

Received 14 February; accepted 10 May 2007.

1. Pounds, J. A. *et al. Nature* **439,** 161–167 (2006).
2. Kiesecker, J. M., Blaustein, A. R. & Belden, L. K. *Nature* **410,** 681–684 (2001).
3. Bosch, J., Carrascal, L. M., Durán, L., Walker, S. & Fisher, M. C. *Proc. R. Soc. B* **274,** 253–260 (2007).
4. Reading, C. J. *Oecologia* **151,** 125–131 (2007).
5. Pascolini, R. *et al. Dis. Aquat. Org.* **56,** 65–74 (2003).
6. Simoncelli, F. *et al. Ecohealth* **2,** 307–312 (2005).
7. Berger, L., Hyatt, A. D., Speare, R. & Longcore, J. E. *Dis. Aquat. Org.* **68,** 51–63 (2005).
8. Stott, P. A., Stone, D. A. & Allen, M. R. *Nature* **432,** 610–614 (2004).
9. Pereira, C. N. *et al. J. Clin. Microbiol.* **43,** 192–198 (2005).
10. Fagotti, A. *et al. Amphibia-Reptilia* **26,** 93–104 (2005).
11. Daszack, P. *et al. Ecology* **86,** 3232–3237 (2005).
12. Lips, K. R. *et al. Proc. Natl Acad. Sci. USA* **103,** 3165–3170 (2006).
13. Garner, T. W. *et al. Emerg. Infect. Dis.* **11,** 1639–1641 (2005).
14. Rachowicz, L. J. *et al. Ecology* **87,** 1671–1683 (2006).
15. Mendelson, J. R. 3rd *et al. Rev. Biol. Trop.* **52,** 991–1000 (2004).

Competing financial interests: declared none.
doi:10.1038/nature05941

ECOLOGY

Pounds *et al.* reply

Replying to: R. A. Alford, K. S. Bradfield & S. J. Richards *Nature* 447, doi:10.1038/nature05940 (2007); I. Di Rosa, F. Simoncelli, A. Fagotti & R. Pascolini *Nature* 447, doi:10.1038/nature05941 (2007)

Alford *et al.*[1] question the working model underlying our test[2] for a link between global warming and amphibian disappearances, and Di Rosa *et al.*[3] criticize our emphasis on a single proximate agent, the chytrid fungus *Batrachochytrium dendrobatidis*. Both teams report key pieces of the amphibian-decline puzzle and new evidence from different parts of the world that climate change is a factor in these losses. Here we show why our working model was appropriate and highlight the complexity of the imminent threat to species survival that results as global warming conspires with various other agents.

We agree with Alford *et al.* that multiyear warm periods are probably more important biologically than single warm years[1]. In the case of the harlequin frogs, *Atelopus*, this may explain why global warming, but not El Niño alone, loads the dice for extinction[2]. Replotting the data from our original Fig. 3a (ref. 2) shows that the warm year just before a disappearance typically followed a series of warm years (Fig. 1). As in Australia[1], raised temperatures characterize the overall period of declines. We found that about 80% of the die-offs corresponded to the warmer years within this period, indicating that the short-term climatic fluctuations probably acted together with the long-term change.

Other studies implicate extremes of climate and the chytrid fungus in these disappearances[4–6], but we agree that various pathogens could be important in such losses[1,3]. In Ecuador, *Telmatobius* frogs showed a high incidence of diseases other than chytridiomycosis during declines, coinciding with warm, dry years[7]. In Venezuela, *Atelopus* and *Nephelobates* frogs showed similar patterns, involving parasitic nematodes and acanthocephalans (J. Piñero and E.L.M., unpublished results)[6]. The rogues' gallery of amphibian pathogens that benefit from climate change is only just beginning to be assembled[3,6–9], and even the chytrid is poorly known[3]. Accordingly, our test of the climate-linked epidemic hypothesis assumed no particular disease or mechanism of outbreak[2].

The potential mechanism of chytridiomycosis outbreaks that we outlined afterwards[2,10] is one of various potential pathways of climate effects that need testing. These chains of events are not mutually exclusive, and their relative importance may vary in space and time[9]. Biotic links might range from a shift in microbial species interactions to a disruption of amphibian development[1–15]. Pathogens may not always be involved[12,15], and there may be other victims, such as reptiles[9,11,15]. We agree that shifts in temperature, precipitation, cloud cover, humidity or winds may be stressors that

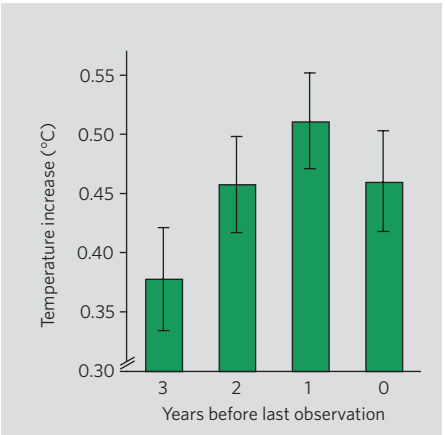


Figure 1 | Punctuated warm periods preceding the disappearances of harlequin frog species (*Atelopus*) in Central and South America. Each bar represents the mean surface temperature for the tropics (departure from a baseline mean for 1856–95) averaged for the set of disappearances ($n = 29$). The bars correspond to the years leading up to disappearances, with the 'zero' bar indicating the last year in which a species was observed. Error bars are 95% confidence intervals. Temperatures typically peaked just before disappearances, but were significantly increased in all of the years represented (see Pounds *et al.*[2] for details).

raise susceptibility to disease[1,3]. New evidence that global warming can degrade toads' body condition and survivorship is compelling in this regard[14]. However, the effects of climate change on host–parasite ecology, and on ecology in general, are manifold.

Global warming also conspires with other forms of environmental deterioration, such as landscape alteration, pollution and species invasions, to dim the prospects for amphibian survival[3,8–10]. For example, humans may have introduced the chytrid into new regions[13]. Because of the presence of exotic species, climate change is an even greater threat. It exerts its influence, not on primeval dioramas, but on present-day ecosystems. It is therefore pointless to debate whether its effects would be noteworthy today if it were acting alone, or whether an agent that it abets would be a problem without it. Likewise, it is fallacious to argue that species that weathered past climate change should survive the present rise in temperatures.

Alford *et al.*[1] and Di Rosa *et al.*[3] add to the mounting evidence that global warming contributes to the present amphibian crisis[1–15], and we reaffirm our test for such a link. The losses portend a planetary-scale mass extinction in the making[10]. Unless humanity takes immediate action to stabilize the climate, while also fighting biodiversity's other threats, a multitude of species is likely to vanish.

J. Alan Pounds*, Martín R. Bustamante†, Luis A. Coloma†, Jamie A. Consuegra‡, Michael P. L. Fogden*, Pru N. Foster§, Enrique La Marca||, Karen L. Masters¶, Andrés Merino-Viteri†, Robert Puschendorf#, Santiago R. Ron†, G. Arturo Sánchez-Azofeifa††, Christopher J. Still‡‡, Bruce E. Young§§**

*Golden Toad Laboratory for Conservation, Monteverde Cloud Forest Preserve and Tropical Science Center, Santa Elena, Puntarenas 5655-73, Costa Rica
e-mail: goldtoad@racsa.co.cr
†Museo de Zoología, Centro de Biodiversidad y Ambiente, Escuela de Biología, Pontificia Universidad Católica del Ecuador, Apartado 17-01-2184, Quito, Ecuador
‡Department of Environmental Science, Barnard College, Columbia University, New York, New York 10027, USA
§Department of Earth Sciences, University of Bristol, Wills Memorial Building, Bristol BS8 1RJ, UK
||Laboratorio de Biogeografía, Escuela de Geografía, Facultad de Ciencias Forestales y Ambientales, Universidad de Los Andes, Apartado 116, Mérida 5101-A, Venezuela
¶Council for International Educational Exchange, Monteverde, Puntarenas 5655-26, Costa Rica
#School of Marine and Tropical Biology, James Cook University, Townsville, Queensland 4811, Australia
**Texas Memorial Museum and Department of Integrative Biology, University of Texas, Austin, Texas 78712, USA
††Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, Alberta T6G 2E3, Canada
‡‡Department of Geography, University of California at Santa Barbara, Santa Barbara, California 93106, USA
§§NatureServe, Plaza Colonial, San José 1260-358, Costa Rica

1. Alford, R. A., Bradfield, K. S. & Richards, S. J. *Nature* **447,** doi:10.1038/nature05940 (2007).
2. Pounds, J. A. *et al. Nature* **439,** 161–167 (2006).
3. Di Rosa, I., Simoncelli, F., Fagotti, A. & Pascolini, R. *Nature* **447,** doi:10.1038/ nature05941 (2007).
4. Ron, S. R., Duellman, W. E., Coloma, L. A. & Bustamante, M. *J. Herpetol.* **37,** 116–126 (2003).
5. Lampo, M., Rodríquez-Contreras, A., La Marca, E. & Daszak, P. *Herpetol. J.* **16,** 395–402 (2006).
6. García, I. J., Albornoz, R. & La Marca, E. *Herpetotropicos* **2,** 64–72 (2007).
7. Merino-Viteri, A., Coloma, L. A. & Almendáriz, A. in *Studies on the Andean Frogs of the Genera Telmatobius and Batrachophrynus* (eds Lavilla, E. O. & De la Riva, I.) 9–37 (Asociación Herpetológica Española, Valencia, 2005).
8. Kiesecker, J. M., Blaustein, A. R. & Belden, L. K. *Nature* **410,** 681–684 (2001).
9. Pounds, J. A. *Nature* **410,** 639–640 (2001).
10. Pounds, J. A. & Puschendorf, R. *Nature* **427,** 107–109 (2004).
11. Pounds, J. A., Fogden, M. P. L. & Campbell, J. H. *Nature* **398,** 611–615 (1999).
12. Daszak, P. *et al. Ecology* **86,** 3232–3237 (2005).
13. Bosch, J., Carrascal, L. M., Durán, L., Walker, S. & Fisher, M. C. *Proc. R. Soc. B* **274,** 253–260 (2007).
14. Reading, C. J. *Oecologia* **151,** 125–131 (2007).
15. Whitfield, S. M. *et al. Proc. Natl Acad. Sci. USA* **104,** 8352–8356 (2007).

doi:10.1038/nature05942

raise susceptibility to disease[1,3]. New evidence that global warming can degrade toads' body condition and survivorship is compelling in this regard[14]. However, the effects of climate change on host–parasite ecology, and on ecology in general, are manifold.

Global warming also conspires with other forms of environmental deterioration, such as landscape alteration, pollution and species invasions, to dim the prospects for amphibian survival[3,8–10]. For example, humans may have introduced the chytrid into new regions[13]. Because of the presence of exotic species, climate change is an even greater threat. It exerts its influence, not on primeval dioramas, but on present-day ecosystems. It is therefore pointless to debate whether its effects would be noteworthy today if it were acting alone, or whether an agent that it abets would be a problem without it. Likewise, it is fallacious to argue that species that weathered past climate change should survive the present rise in temperatures.

Alford *et al.*[1] and Di Rosa *et al.*[3] add to the mounting evidence that global warming contributes to the present amphibian crisis[1–15], and we reaffirm our test for such a link. The losses portend a planetary-scale mass extinction in the making[10]. Unless humanity takes immediate action to stabilize the climate, while also fighting biodiversity's other threats, a multitude of species is likely to vanish.

J. Alan Pounds*, Martín R. Bustamante†, Luis A. Coloma†, Jamie A. Consuegra‡, Michael P. L. Fogden*, Pru N. Foster§, Enrique La Marca||, Karen L. Masters¶, Andrés Merino-Viteri†, Robert Puschendorf#, Santiago R. Ron†, G. Arturo Sánchez-Azofeifa††, Christopher J. Still‡‡, Bruce E. Young§§**

*Golden Toad Laboratory for Conservation, Monteverde Cloud Forest Preserve and Tropical Science Center, Santa Elena, Puntarenas 5655-73, Costa Rica
e-mail: goldtoad@racsa.co.cr
†Museo de Zoología, Centro de Biodiversidad y Ambiente, Escuela de Biología, Pontificia Universidad Católica del Ecuador, Apartado 17-01-2184, Quito, Ecuador
‡Department of Environmental Science, Barnard College, Columbia University, New York, New York 10027, USA
§Department of Earth Sciences, University of Bristol, Wills Memorial Building, Bristol BS8 1RJ, UK
||Laboratorio de Biogeografía, Escuela de Geografía, Facultad de Ciencias Forestales y Ambientales, Universidad de Los Andes, Apartado 116, Mérida 5101-A, Venezuela
¶Council for International Educational Exchange, Monteverde, Puntarenas 5655-26, Costa Rica
#School of Marine and Tropical Biology, James Cook University, Townsville, Queensland 4811, Australia
**Texas Memorial Museum and Department of Integrative Biology, University of Texas, Austin, Texas 78712, USA
††Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, Alberta T6G 2E3, Canada
‡‡Department of Geography, University of California at Santa Barbara, Santa Barbara, California 93106, USA
§§NatureServe, Plaza Colonial, San José 1260-358, Costa Rica

1. Alford, R. A., Bradfield, K. S. & Richards, S. J. *Nature* **447,** doi:10.1038/nature05940 (2007).
2. Pounds, J. A. *et al. Nature* **439,** 161–167 (2006).
3. Di Rosa, I., Simoncelli, F., Fagotti, A. & Pascolini, R. *Nature* **447,** doi:10.1038/ nature05941 (2007).
4. Ron, S. R., Duellman, W. E., Coloma, L. A. & Bustamante, M. *J. Herpetol.* **37,** 116–126 (2003).
5. Lampo, M., Rodríquez-Contreras, A., La Marca, E. & Daszak, P. *Herpetol. J.* **16,** 395–402 (2006).
6. García, I. J., Albornoz, R. & La Marca, E. *Herpetotropicos* **2,** 64–72 (2007).
7. Merino-Viteri, A., Coloma, L. A. & Almendáriz, A. in *Studies on the Andean Frogs of the Genera Telmatobius and Batrachophrynus* (eds Lavilla, E. O. & De la Riva, I.) 9–37 (Asociación Herpetológica Española, Valencia, 2005).
8. Kiesecker, J. M., Blaustein, A. R. & Belden, L. K. *Nature* **410,** 681–684 (2001).
9. Pounds, J. A. *Nature* **410,** 639–640 (2001).
10. Pounds, J. A. & Puschendorf, R. *Nature* **427,** 107–109 (2004).
11. Pounds, J. A., Fogden, M. P. L. & Campbell, J. H. *Nature* **398,** 611–615 (1999).
12. Daszak, P. *et al. Ecology* **86,** 3232–3237 (2005).
13. Bosch, J., Carrascal, L. M., Durán, L., Walker, S. & Fisher, M. C. *Proc. R. Soc. B* **274,** 253–260 (2007).
14. Reading, C. J. *Oecologia* **151,** 125–131 (2007).
15. Whitfield, S. M. *et al. Proc. Natl Acad. Sci. USA* **104,** 8352–8356 (2007).

doi:10.1038/nature05942

ARTICLES

Two neurons mediate diet-restriction-induced longevity in *C. elegans*

Nicholas A. Bishop[1] & Leonard Guarente[1]

Dietary restriction extends lifespan and retards age-related disease in many species and profoundly alters endocrine function in mammals. However, no causal role of any hormonal signal in diet-restricted longevity has been demonstrated. Here we show that increased longevity of diet-restricted *Caenorhabditis elegans* requires the transcription factor gene *skn-1* acting in the ASIs, a pair of neurons in the head. Dietary restriction activates *skn-1* in these two neurons, which signals peripheral tissues to increase metabolic activity. These findings demonstrate that increased lifespan in a diet-restricted metazoan depends on cell non-autonomous signalling from central neuronal cells to non-neuronal body tissues, and suggest that the ASI neurons mediate diet-restriction-induced longevity by an endocrine mechanism.

Dietary restriction extends lifespan in many organisms and reduces incidence and progression of age-related disease[1]. In mammals, dietary restriction dramatically alters central and secondary hormone production and target-tissue responsiveness, suggesting a mechanism by which this regimen coordinately slows ageing across a variety of tissues[2]. A possible hormonal link of dietary restriction to longevity is suggested by the recent demonstration that serum from diet-restricted mice contains factors sufficient to induce many of the cellular phenotypes of dietary restriction *in vitro*[3]. However, no hormone has been shown to mediate diet-restriction-induced longevity *in vivo*, and indeed mice that almost completely lack several pituitary hormones respond normally to dietary restriction[4], so the question of whether dietary restriction induces longevity by an endocrine mechanism remains unresolved.

A method of dietary restriction in *C. elegans*

Dietary restriction has been shown to extend lifespan in the nematode worm *C. elegans*[5]. We developed a dietary restriction protocol that reliably increases longevity in worms (see Methods; Fig. 1a). Worms cultured on standard plates seeded with the bacterial food source throughout larval development were then transferred to a small volume of liquid medium in which bacterial density was controlled at a range of different levels (see Methods). Worms maintained at high bacterial density (our *ad libitum* condition) had a slightly longer lifespan than worms cultured on agar plates seeded with a bacterial lawn (Fig. 1a). Mean lifespan increased as bacterial density was reduced until a maximum lifespan was reached at an optimal dilution of 1:10 (our dietary restriction condition), although further dilutions caused reduced mean lifespan, presumably due to starvation (Fig. 1b). Diet-restricted animals also exhibited reduced fat in the intestine and small body size (Supplementary Fig. 1a, b). Animals in the *ad libitum* condition exhibited an optimal brood size, whereas diet-restricted animals had a reduced brood size and extended reproductive period (Supplementary Fig. 1c). As reported for other methods of dietary restriction[6,7], this protocol extended the lifespan of both *daf-16* and *daf-2* mutants (Fig. 1c) disrupted in genes that function in the well-characterized insulin-like signalling pathway[8].

SKN-1 mediates diet-restricted longevity

SKN-1 is a transcription factor that has previously been shown to be critical in endodermal development[9] and control of the oxidative stress response[10], functions that are shared by the most similar mammalian proteins, the NFE2-related factors. Because oxidative damage is thought to be a key cause of ageing[11], it has been proposed that SKN-1 may be a determinant of ageing rate[10]. Notably, some isoforms of SKN-1 are expressed from an operon downstream of *bec-1*, the *C. elegans* homologue of mammalian beclin 1 (Fig. 2a). Beclin 1 mediates autophagy induced by nutrient deprivation[12], indicating that *skn-1* might be regulated in response to nutritional stress. We therefore tested whether *skn-1* functions in the diet-restriction-induced longevity response. Four loss-of-function alleles of *skn-1* all significantly impaired the response to dietary restriction (Fig. 2b; Supplementary Figs 2a and 3). *skn-1* mutant lifespan was unresponsive to diet over a wide range of food concentrations (Fig. 1b). *skn-1* mutations seemed to specifically interfere with the dietary restriction longevity response, because they did not impair the longevity increase caused by mutation of *daf-2* (Supplementary Fig. 4) or by treatment with a plant polyphenol extract[13]. We focused our subsequent analysis on two alleles, *skn-1(zu135)* and *skn-1(zu169)*, because they completely prevented the longevity increase on dietary restriction, yet had little or no effect on lifespan of worms fed *ad libitum* (Fig. 2b). These two alleles introduce premature stop codons before the sequence encoding the DNA-binding domain of all SKN-1 isoforms (Supplementary Fig 2a; also see below). The failure of both *skn-1* mutants to respond to dietary restriction was rescued by *Is007*, an integrated transgene expressing an *skn-1–gfp* (green fluorescent protein) fusion from its native promoter[10] (Fig. 2a, c; Supplementary Fig. 2a), confirming that these *skn-1* mutations caused the dietary restriction longevity defect.

skn-1 mediates dietary restriction in two neurons

The rescuing *Is007* transgene is expressed in only two tissues: in the two ASI neurons, and in the intestine[10]. This is an intriguing expression pattern in the context of dietary restriction because the intestine is the food absorption and storage organ, and the ASIs are sensory neurons in the head that are critical in translating information about food availability into endocrine signals that influence the dauer decision during development[14]. We asked in which tissue *skn-1* acts to mediate dietary restriction longevity. The *skn-1* gene encodes three protein isoforms—SKN-1A, SKN-1B and SKN-1C—which have different amino termini but a common carboxy terminus (Fig. 2a).

[1]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Using 5′ rapid amplification of cloned/cDNA ends (RACE), we obtained evidence that each of these isoforms is expressed from an independent promoter (Supplementary Fig. 2b–d). Because most of the unique 5′ coding region of *skn-1a* is not included in the rescuing *Is007* transgene (Fig. 2a), we reasoned that *skn-1b* and/or *skn-1c* must be sufficient to rescue dietary restriction defects in the *skn-1* mutants. We used RNA interference (RNAi) feeding clones corresponding to the unique 5′ end of either *skn-1b* or *skn-1c* to knock down each isoform separately in GFP reporter strains, and determined that *skn-1b* is expressed in the ASI neurons but not detectably in the gut, and *skn-1c* is expressed in the gut but not detectably in the ASI neurons (Fig. 2d–f). An extrachromosomal transgene that drives *skn-1b–gfp* expression from the ASI-specific *gpa-4* promoter[15] showed expression in the ASI neurons, but not the gut, and rescued the dietary restriction longevity defects of *skn-1(zu135)* animals completely (Fig. 2g). Similar results were obtained with the same transgene in a *skn-1(zu169)* background, and also with a chromosomally integrated version of the transgene (Supplementary Table 1). In contrast, extrachomosomal and integrated transgenes driving *skn-1c–gfp* from the *ges-1* promoter[16] showed expression only in the gut, and had no effect on the dietary restriction defect of *skn-1(zu135)* (Fig. 2h). Furthermore, ectopic expression of *skn-1c–gfp* in the ASI neurons using the *gpa-4* promoter failed to rescue the dietary restriction longevity defect of *skn-1(zu135)* (Supplementary Table 1). Taken together, these data indicate that *skn-1b* functions in the ASI neurons to mediate dietary restriction longevity.

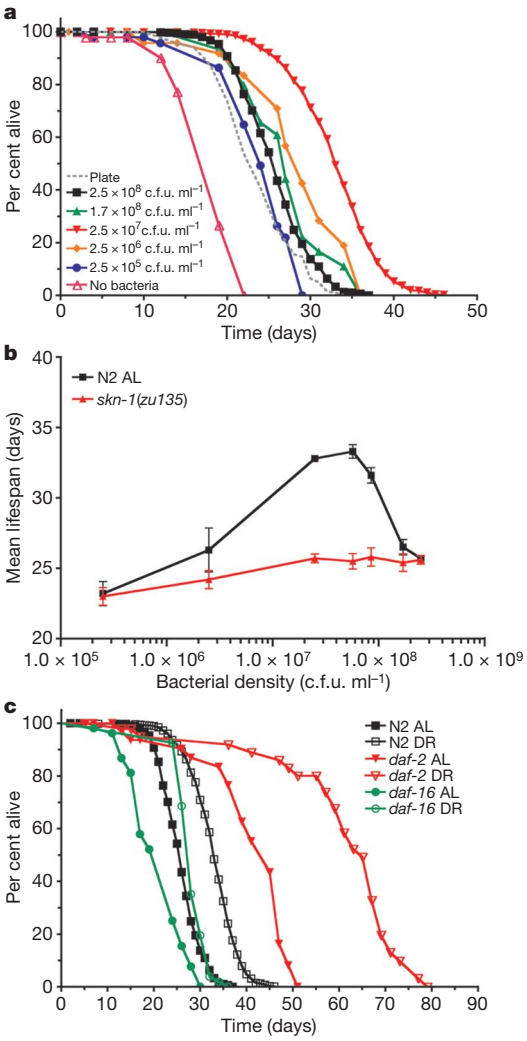


Figure 1 | Lifespan extension by dietary restriction. Lifespan curves represent combined data from independent experiments as indicated in Supplementary Table 1. Complete lifespan data are presented in Supplementary Table 1. **a**, Survival curves of wild-type N2 animals fed various concentrations of bacteria in liquid medium, or on agar plates of otherwise identical composition seeded with a bacterial lawn. **b**, Local maximum in N2 mean lifespan at an optimal level of bacterial dilution. In contrast, the lifespan of *skn-1(zu135)* is not altered by food level. Error bars, s.e.m. **c**, The dietary restriction (DR) protocol extends lifespan of the insulin pathway mutants *daf-16(mgDf50)* and *daf-2(e1370)*. *Ad libitum* (AL) bacterial concentration is 2.5×10^8 c.f.u. ml^{-1}; dietary restriction is 2.5×10^7 c.f.u. ml^{-1}.

To confirm that the ASI neurons are required for diet-restriction-induced longevity, we used a laser microbeam to specifically kill these two cells and tested the effect in subsequent dietary restriction. Ablation of the ASI neurons completely suppressed the response to dietary restriction (Fig. 3a), but also caused a small increase in basal longevity, consistent with a previous report[17]. The ASIs can apparently affect lifespan by two independent pathways because the increase in basal longevity was previously reported to be dependent on *daf-16* (ref. 17), and dietary restriction longevity is not (Fig. 1c). We reasoned that the use of a *daf-16* mutant would separate the two effects of ablating the ASI neurons and allow us to determine their role in dietary restriction longevity specifically. As expected, when we repeated the ASI ablation in *daf-16* animals, we found that basal longevity was unaffected by the ablation (Fig. 3b). Importantly, dietary restriction longevity was still suppressed (Fig. 3b). These results support the model that *skn-1* functions in the ASI neurons to extend lifespan in response to dietary restriction.

We next investigated how *skn-1* acts in the ASI to mediate dietary restriction longevity. To rule out the possibility that *skn-1* mutation disrupts development of the ASI neurons, we used a panel of ASI-specific GFP reporters to confirm that the ASI neurons exhibit normal morphology and cell fate in a *skn-1* mutant background (Supplementary Fig. 5). We then tested whether SKN-1–GFP is induced by dietary restriction in *Is007* animals. Although we saw no specific induction in the intestine (data not shown), dietary restriction significantly increased SKN-1–GFP fluorescence in the ASI neurons (Fig. 4a). The fluorescence increase was specific to SKN-1–GFP expressed from its native promoter, because if the protein was expressed from the ASI-specific *gpa-4* promoter, we observed instead a slight decrease in expression during dietary restriction (Fig. 4a). This dependence of *skn-1* induction on the native promoter indicates that *skn-1* may be transcriptionally regulated during dietary restriction.

The ASI neurons activate respiration

In yeast, dietary restriction induces an increase in respiratory rate[18], and there is evidence that this is also the case in worms[19,20]. We measured the whole-body oxygen consumption rate of large populations of wild-type worms on day 3 of *ad libitum* feeding or dietary restriction (see Methods), and found that diet-restricted worms exhibited elevated respiration (Fig. 4b). The respiration rate increase on dietary restriction was absent in a *skn-1(zu135)* mutant, but could be rescued, and somewhat enhanced, by the *Is007* transgene (Fig. 4b), demonstrating that *skn-1* is necessary for the increased respiration. The increase in respiration is likely to be necessary for the dietary restriction longevity effect because two different specific inhibitors of the mitochondrial electron transport chain complex III, myxothiazol and antimycin, suppressed dietary restriction longevity without shortening the lifespan of worms fed *ad libitum* (Fig. 4c, d). Furthermore, myxothiazol completely suppressed the increase in respiration under dietary restriction (data not shown). The effect of inhibiting electron transport was specific to dietary restriction longevity because the long life of a *daf-2* mutant was not affected by antimycin (Fig. 4e). Finally, to determine whether *skn-1* in the ASI neurons alone is sufficient to rescue the absence of increased respiration in the *skn-1(zu135)* mutant during dietary restriction, we constructed two stably integrated transgenes: *geIs9*, which drives

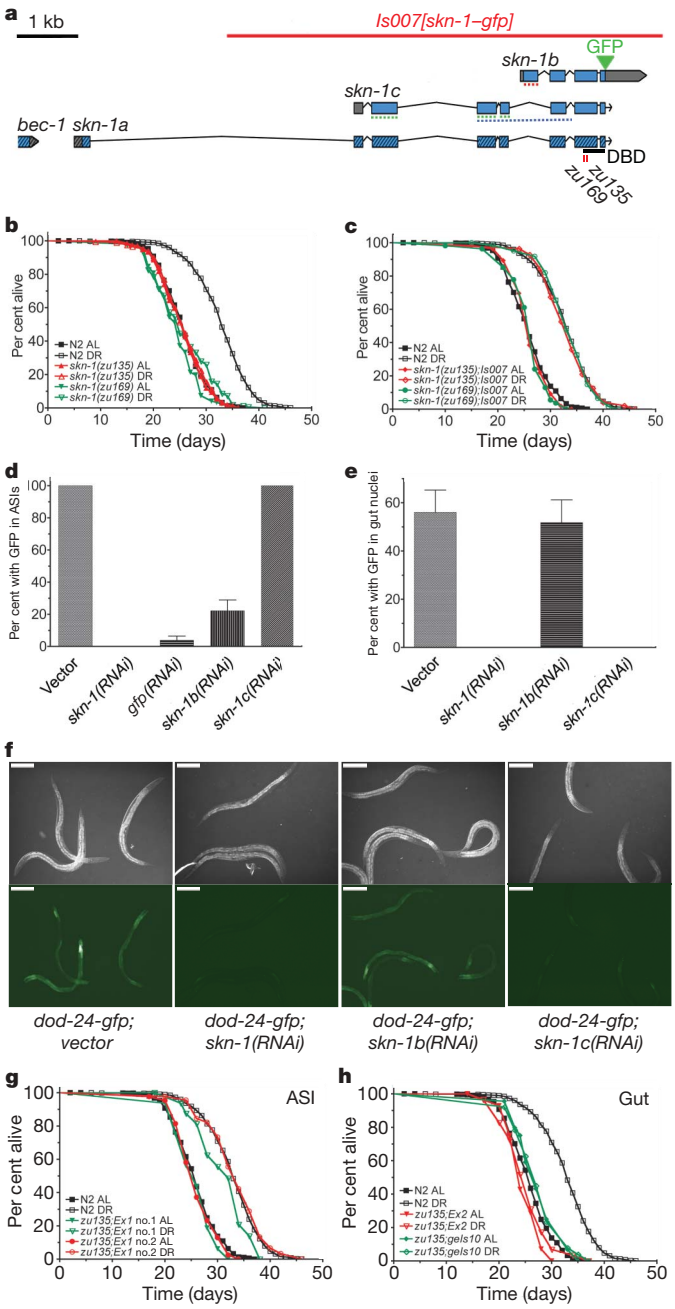


Figure 2 | *skn-1* functions in the ASI neurons to mediate dietary restriction longevity. **a**, The *skn-1* gene encodes three protein isoforms, with different N termini but a common C terminus. Predicted transcripts (Wormbase) and their genomic organization are depicted, with exons shown in blue and untranslated regions in grey. Striped boxes indicate an operonic transcript including the upstream gene—the beclin-1 homologue *bec-1*—and downstream *skn-1a*. The extent of the *Is007[skn-1–gfp]* integrated transgene[10] and the insertion site of GFP in the transgene are also indicated. The purple and red dotted lines indicate the DNA segments used in the *skn-1(RNAi)* and *skn-1b(RNAi)* constructs, respectively (see below); the *skn-1c(RNAi)* construct was made using a fusion of the exons indicated by the green dotted line. The sites of the *zu169* and *zu135* nonsense mutations are shown. DBD, DNA binding domain. See Supplementary Fig. 2a for more detail. **b**, Two strains carrying *skn-1* mutations *zu135* and *zu169* have normal basal lifespans, but fail to respond to dietary restriction. **c**, Both *skn-1* mutants are rescued by the integrated transgene *Is007*, which drives *skn-1–gfp* from its native promoter and expresses *skn-1–gfp* in the ASI neurons and the gut. **d–f**, The *skn-1b* isoform is expressed primarily in the ASI neuron, and *skn-1c* is expressed primarily in the intestine. RNAi constructs specific to the unique 5′ end of either *skn-1b* or *skn-1c* (see panel **a**) were used to examine *skn-1* isoform expression. Note that the *skn-1c* RNAi construct will also knock down endogenous *skn-1a* owing to sequence identity in this region. The *skn-1(RNAi)* construct targets the 3′ region common to all *skn-1* isoforms (see panel **a**). **d**, *Is007* expresses SKN-1B–GFP, but not SKN-1C–GFP, in the ASI. The indicated RNAi constructs were fed to RNAi-hypersensitive[30] *lin-35(n745);Is007[skn-1–gfp]* animals. *skn-1b(RNAi)*, but not *skn-1c(RNAi)*, reduced GFP expression in the ASI. Data shown are from two independent experiments; errors are s.e.m. Total *n* of animals observed in trial 1 and trial 2, respectively: vector, 24, 29; *skn-1(RNAi)*, 20, 24; *gfp(RNAi)*, 22, 26; *skn-1b(RNAi)*, 36, 24; *skn-1c(RNAi)*, 23,20. **e**, *Is007* expresses SKN-1C–GFP, but not SKN-1B–GFP, in the intestine. Following RNAi exposure, *Is007* worms were heat-shocked to induce nuclear accumulation of SKN-1–GFP in the intestine. *skn-1c(RNAi)* reduced SKN-1–GFP, whereas *skn-1b(RNAi)* had no effect. Data shown are from two independent experiments; errors are s.e.m. Total *n* of animals examined in trial 1 and trial 2, respectively: vector, 24, 25; *skn-1(RNAi)*, 25, 23; *skn-1b(RNAi)*, 28, 24; *skn-1c(RNAi)*, 28, 21. **f**, *skn-1a* and/or *skn-1c* account for the majority of basal *skn-1* transcriptional activity in the intestine. The gene *dod-24* has several *skn-1* consensus binding sites in its promoter region and is expressed in a *skn-1*-dependent manner (data not shown). An integrated *dod-24-gfp* transgenic strain (a gift of D. Kim and F. Ausubel) was therefore used as a reporter for intestinal *skn-1* activity. *skn-1c(RNAi)* reduced reporter expression, whereas *skn-1b(RNAi)* did not. Scale bars, 100 μm. **g**, *Ex1*, an extrachromosomal array that drives *skn-1b–gfp* expression from the *gpa-4* promoter exclusively in the ASI neurons of adult animals, rescues the *skn-1(zu135)* dietary restriction longevity defect. Two independently transformed lines, labelled no. 1 and no. 2, are shown. **h**, *Ex2*, an extrachromosomal array driving *skn-1c–gfp* expression from the *ges-1* promoter exclusively in the gut, does not rescue *skn-1(zu135)* dietary restriction longevity. *geIs10*, an integrated version of the same array, also fails to rescue.

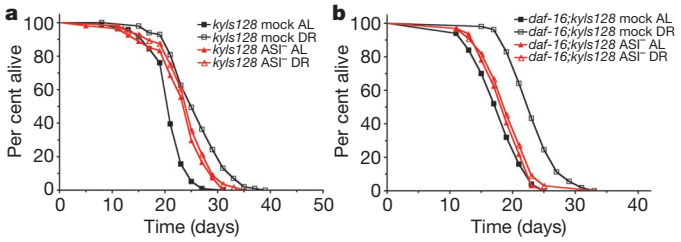


Figure 3 | The ASI neurons are necessary for dietary restriction longevity. ASIs were laser-ablated in strains carrying the integrated array *kyIs128*, which expresses GFP from the ASI-specific *str-3* promoter[27]. 'Mock' strains represent non-operated controls (see Methods). **a**, Animals with killed ASIs (ASI^-) do not respond to dietary restriction. These animals have a longer basal lifespan, as previously reported[17] (for *kyIs128* ASI^- AL versus *kyIs128* mock AL, Mantel–Cox logrank $P < 0.0001$). **b**, *daf-16* mutation suppresses the longer basal lifespan caused by ASI ablation (for *daf-16(mgDf50);kyIs128* ASI^- AL versus *daf-16(mgDf50);kyIs128* mock AL, $P = 0.3902$), but ASI-ablated animals remain refractory to dietary restriction.

skn-1b–gfp expression in the ASI neurons, and *geIs10*, which drives *skn-1c–gfp* expression in the intestine. *geIs9* expressed strongly in the ASIs, and also rescued and enhanced the respiration response to dietary restriction in the *skn-1(zu135)* mutant (Fig. 4b). The *geIs9* line sometimes showed weak expression in a few cells apart from the ASIs, although these are unlikely to contribute to the observed rescue (see Methods). *geIs10*, in contrast, failed to rescue the dietary restriction respiration defect of *skn-1(zu135)*, despite expressing efficiently in the intestine (Supplementary Fig. 6). These findings suggest that diet-restriction-induced activation of *skn-1* in the ASIs causes release of a signal that promotes metabolism in peripheral tissues and leads to long life. Interestingly, it seems that *skn-1* has two functions in the adult that are separable by isoform and tissue of expression: mediation of dietary restriction longevity by *skn-1b* in the ASI neurons,

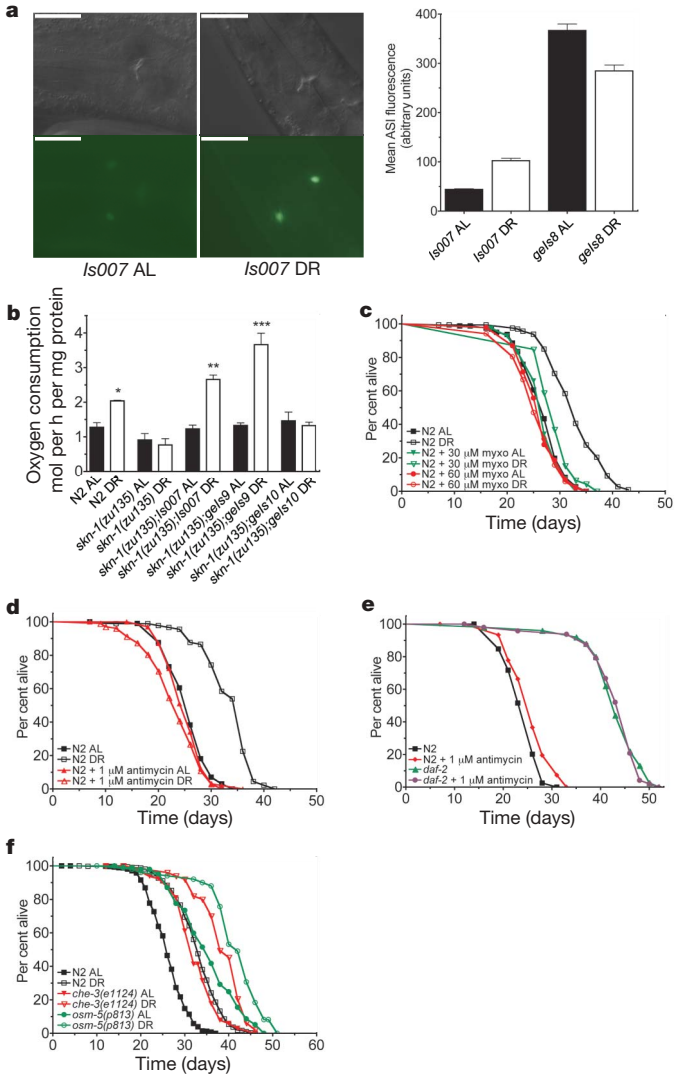


Figure 4 | Dietary restriction activates *skn-1* in the ASI neurons and increases whole-body respiration. **a**, Five days of dietary restriction increases GFP expression in the ASI neurons of *Is007[skn-1–gfp]* worms. Left panel, bright field and fluorescence Nomarski images; scale bars, 10 μm. Right panel, quantification of fluorescence in *Is007[skn-1–gfp]* and in *geIs8*, a specificity control strain that drives *skn-1–gfp* expression from the ASI-specific *gpa-4* promoter. Each value represents pooled data from two independent trials; errors are s.e.m.; *n* for trial 1 and trial 2, respectively: *Is007* AL, 42, 42; *Is007* DR, 58, 21 (unpaired two-tailed *t*-test $P < 0.0001$); *geIs8* AL, 19, 31; *geIs8* DR, 19, 39 ($P < 0.0001$). **b**, Respiration rate is increased in N2 animals assayed on day 3 of dietary restriction. The *skn-1(zu135)* mutation prevents the respiration increase on dietary restriction. Integrated transgenes driving *skn-1–gfp* expression, either from the native promoter (*Is007*), or from the ASI-specific *gpa-4* promoter (*geIs9*), rescue the dietary restriction respiration defect of *skn-1(zu135)*, elevating respiration rates above the wild-type level. The integrated transgene *geIs10*, which drives *skn-1–gfp* expression in the gut using the *ges-1* promoter, fails to rescue the respiration defect of *skn-1(zu135)* on dietary restriction. Unpaired two tailed *t*-test *P* values (AL versus DR for each strain): N2, $*P = 0.0221$; *skn-1(zu135);Is007*, $**P = 0.0011$; *skn-1(zu135);geIs9*, $***P = 0.0005$. No other values differ significantly from N2 AL. N2 data represent pooled triplicate measurements from each of four independent experiments. Other strains were measured in triplicate in two experiments each; errors are s.e.m. **c**, **d**, Electron-transport-chain-inhibiting drugs prevent the dietary restriction longevity response without shortening AL lifespan. The electron transport chain complex III inhibitor myxothiazol (myxo) partially blocks dietary restriction longevity at 30 μM, and completely blocks dietary restriction longevity at 60 μM (**c**). Another complex III inhibitor, antimycin, also blocks dietary restriction longevity (**d**). **e**, Antimycin does not affect the longevity induced by *daf-2(e1370)* mutation. This assay was performed on agar plates with otherwise identical compositions to the liquid medium used for dietary restriction experiments. **f**, The dietary restriction longevity response does not require environmental chemosensation. *che-3(e1124)* and *osm-5(p813)* mutants lack functional sensory neuron cilia and are long-lived under AL conditions[31], but dietary restriction still increases lifespan in these mutants by a similar percentage as it does in N2.

and resistance to oxidative stress by *skn-1c* in the intestine (Supplementary Fig. 6).

Discussion

How might dietary restriction activate *skn-1* in the ASIs? Although the ASI is a sensory neuron[21], sensation of the environment by the ASI is unlikely to be necessary for dietary restriction to increase longevity because mutants lacking functional sensory cilia still expressed SKN-1–GFP in the ASI neurons (not shown) and responded normally to dietary restriction (Fig. 4f). It has recently been shown that, within particular cells of the mammalian hypothalamus, levels of key intracellular metabolites indicative of energy availability control organismal food intake and energy metabolism[22]. We propose that the ASI neurons similarly detect their own cellular energy state and adjust the animal's overall metabolism in response to nutrient availability. Because increased expression of *skn-1–gfp* in the ASI is not sufficient by itself to extend the lifespan of worms fed *ad libitum* or to enhance the response to dietary restriction (Supplementary Table 1), other pathways or neurons must act in parallel to *skn-1* to orchestrate dietary restriction longevity.

Here we demonstrate that cell non-autonomous signals from central neuroendocrine cells to the periphery can activate respiration and mediate dietary restriction longevity in a metazoan (Fig. 5). The ASIs have well-established endocrine functions, and express at least 15 known or potential hormones, including members of the insulin, TGF-β, and neuropeptide classes[23,24]. We propose that, on sensing dietary restriction, the ASI neurons release, or cause other neurons to release, a hormonal signal that promotes metabolic activity in the periphery and leads to increased respiration and long life. Mutations in genes encoding two hormones produced by the ASI neurons, an insulin (*daf-28*) and a TGF-β (*daf-7*), do not affect the dietary restriction longevity response, suggesting that these genes do not have an essential role in dietary restriction (N.B and L.G., unpublished results). We anticipate that future work will dissect the critical central hormones mediating diet-restriction-induced longevity of worms and mammals.

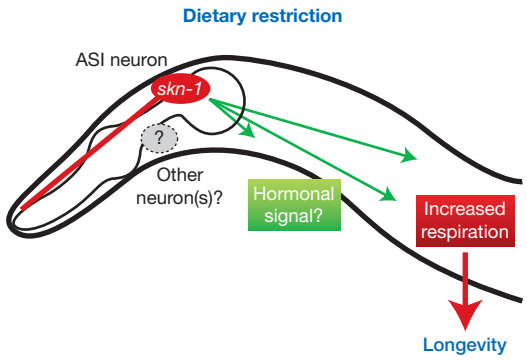


Figure 5 | Model of *skn-1* function during dietary restriction. Dietary restriction activates *skn-1* specifically in the ASI neurons. This promotes cell non-autonomous signalling to the peripheral tissues that increases respiration and extends lifespan. Other neurons or genetic pathways may be required in parallel to *skn-1* to mediate dietary restriction longevity (see text).

METHODS SUMMARY

Dietary restriction protocol. Lifespan assays were performed in 6-well tissue culture plates, with each well containing 2.5 ml of dietary restriction bottom medium and 2.5 ml of dietary restriction top medium. Dietary restriction bottom medium is composed of standard NGM medium[25], supplemented with 1 mg ml^{-1} erythromycin to prevent bacterial division, 12.5 μg ml^{-1} fluorodeoxyuridine to inhibit progeny hatching, 50 μg ml^{-1} ampicillin, and 1 mM isopropyl-β-D-thiogalactoside; dietary restriction top medium is identical to dietary restriction bottom medium, excluding the agar. Ampicillin-resistant bacteria were added from a concentrated stock to the desired concentration, and concentration was monitored using a spectrophotometer. Worms were grown to the L4/young adult stage on NGM plates seeded with OP50 bacteria, then transferred individually into the wells and maintained at 20 °C with gentle gyratory shaking.

Laser ablation. Synchronized L1 animals were ablated, or mock-ablated in parallel, using a laser microbeam as described[26]. Worms were recovered to plates and grown to the L4/young adult stage before lifespan analysis. The *kyIs128* strain[27] proved to have a somewhat shorter lifespan than N2, but this was attributable to intrinsic strain differences and not to the mounting procedure (Supplementary Fig. 7).

Respiration assay. Mass plate cultures of synchronized worms were grown on OP50 plates to the L4 stage, then transferred to Erlenmeyer flasks containing *ad libitum* or dietary restriction levels of bacteria. After three days of culture, animals were collected by centrifugation and washed free of bacteria in S buffer. Respiration rate was measured at 25 °C and normalized to protein content as described[28].

Fluorescence intensity quantification. Fluorescent images were collected at 1,000× magnification (Zeiss) from worms subjected to 5 days of *ad libitum* feeding or dietary restriction, as described above. Fluorescence brightness in the ASI was quantified using NIH ImageJ software.

RNAi feeding. RNAi feeding assays were done as described previously[29].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 February; accepted 9 May 2007.

1. Koubova, J. & Guarente, L. How does calorie restriction work? *Genes Dev.* **17,** 313–321 (2003).
2. Mobbs, C. V. *et al.* Neuroendocrine and pharmacological manipulations to assess how caloric restriction increases life span. *J. Gerontol. A Biol. Sci. Med. Sci.* 56 (Spec No 1) 34–44 (2001).
3. de Cabo, R. *et al.* An *in vitro* model of caloric restriction. *Exp. Gerontol.* **38,** 631–639 (2003).
4. Bartke, A. *et al.* Extending the lifespan of long-lived mice. *Nature* **414,** 412 (2001).
5. Walker, G., Houthoofd, K., Vanfleteren, J. R. & Gems, D. Dietary restriction in *C. elegans*: from rate-of-living effects to nutrient sensing pathways. *Mech. Ageing Dev.* **126,** 929–937 (2005).
6. Lakowski, B. & Hekimi, S. The genetics of caloric restriction in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **95,** 13091–13096 (1998).
7. Houthoofd, K., Braeckman, B. P., Johnson, T. E. & Vanfleteren, J. R. Life extension via dietary restriction is independent of the Ins/IGF-1 signalling pathway in *Caenorhabditis elegans*. *Exp. Gerontol.* **38,** 947–954 (2003).
8. Kenyon, C. The plasticity of aging: insights from long-lived mutants. *Cell* **120,** 449–460 (2005).
9. Bowerman, B., Eaton, B. A. & Priess, J. R. *skn-1*, a maternally expressed gene required to specify the fate of ventral blastomeres in the early *C. elegans* embryo. *Cell* **68,** 1061–1075 (1992).
10. An, J. H. & Blackwell, T. K. SKN-1 links *C. elegans* mesendodermal specification to a conserved oxidative stress response. *Genes Dev.* **17,** 1882–1893 (2003).
11. Golden, T. R., Hinerfeld, D. A. & Melov, S. Oxidative stress and aging: beyond correlation. *Aging Cell* **1,** 117–123 (2002).
12. Zhenyu, Y., Jin, S., Yang, C., Levine, A. J. & Heintz, N. Beclin 1, an autophagy gene essential for early embryonic development, is a haploinsufficient tumor suppressor. *Proc. Natl Acad. Sci. USA* **100,** 15077–15082 (2003).
13. Wilson, M. A. *et al.* Blueberry polyphenols increase lifespan and thermotolerance in *Caenorhabditis elegans*. *Aging Cell* **5,** 59–68 (2006).
14. Bargmann, C. I. & Horvitz, H. R. Control of larval development by chemosensory neurons in *Caenorhabditis elegans*. *Science* **251,** 1243–1246 (1991).
15. Jansen, G. *et al.* The complete family of genes encoding G proteins of *Caenorhabditis elegans*. *Nature Genet.* **21,** 414–419 (1999).
16. Libina, N., Berman, J. R. & Kenyon, C. Tissue-specific activities of *C. elegans* DAF-16 in the regulation of lifespan. *Cell* **115,** 489–502 (2003).
17. Alcedo, J. & Kenyon, C. Regulation of *C. elegans* longevity by specific gustatory and olfactory neurons. *Neuron* **41,** 45–55 (2004).
18. Lin, S. J. *et al.* Calorie restriction extends *Saccharomyces cerevisiae* lifespan by increasing respiration. *Nature* **418,** 344–348 (2002).
19. Houthoofd, K. *et al.* Axenic growth up-regulates mass-specific metabolic rate, stress resistance, and extends life span in *Caenorhabditis elegans*. *Exp. Gerontol.* **37,** 1371–1378 (2002).
20. Houthoofd, K. *et al.* No reduction of metabolic rate in food restricted *Caenorhabditis elegans*. *Exp. Gerontol.* **37,** 1359–1369 (2002).
21. Bargmann, C. I. & Horvitz, H. R. Chemosensory neurons with overlapping functions direct chemotaxis to multiple chemicals in *C. elegans*. *Neuron* **7,** 729–742 (1991).
22. Dowell, P., Hu, Z. & Lane, M. D. Monitoring energy balance: metabolites of fatty acid synthesis as hypothalamic sensors. *Annu. Rev. Biochem.* **74,** 515–534 (2005).
23. Li, C. The ever-expanding neuropeptide gene families in the nematode *Caenorhabditis elegans*. *Parasitology* **131** (Suppl.) S109–S127 (2005).
24. Ren, P. *et al.* Control of *C. elegans* larval development by neuronal expression of a TGF-β homolog. *Science* **274,** 1389–1391 (1996).
25. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77,** 71–94 (1974).
26. Avery, L. & Horvitz, H. R. Pharyngeal pumping continues after laser killing of the pharyngeal nervous system of *C. elegans*. *Neuron* **3,** 473–485 (1989).
27. Peckol, E. L., Zallen, J. A., Yarrow, J. C. & Bargmann, C. I. Sensory activity affects sensory axon development in *C. elegans*. *Development* **126,** 1891–1902 (1999).
28. Braeckman, B. P., Houthoofd, K. & Vanfleteren, J. R. Assessing metabolic activity in aging *Caenorhabditis elegans*: concepts and controversies. *Aging Cell* **1,** 82–88 (2002).
29. Timmons, L., Court, D. L. & Fire, A. Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* **263,** 103–112 (2001).
30. Wang, D. *et al.* Somatic misexpression of germline P granules and enhanced RNA interference in retinoblastoma pathway mutants. *Nature* **436,** 593–597 (2005).
31. Apfeld, J. & Kenyon, C. Regulation of lifespan by sensory perception in *Caenorhabditis elegans*. *Nature* **402,** 804–809 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. R. Horvitz for allowing use of essential equipment, and members of the Guarente and Horvitz laboratories for advice and discussions. We thank D. Kim and F. Ausubel for the gift of an unpublished strain. Many of the strains used in this work were provided by the *Caenorhabditis* Genetics Center. This work was supported by a grant from the National Institutes of Health.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.G. (leng@mit.edu).

METHODS

Strains. We used the following strains: the wild-type N2, EU1 *skn-1(zu67)/ nT1[unc-?(n754);let-?]*, EU31 *skn-1(zu135)/ nT1[unc-?(n754);let-?]*, EU35 *skn-1(zu169)/nT1[unc-?(n754);let-?]*, EU40 *skn-1(zu129)/ nT1[unc-?(n754);let-?]*, LG335 *skn-1(zu135)/nT1[qIs51]*, LG336 *skn-1(zu169)/nT1[qIs51]*, LG333 *skn-1(zu135);Is007[skn-1–gfp]*, LG326 *skn-1(zu169);Is007*, LG340 *skn-1(zu135)/nT1 [qIs51];Ex(gpa-4-skn-1b–gfp)* no. 1, LG341 *skn-1(zu169)/nT1[qIs51];Ex(gpa-4-skn-1b–gfp)* no. 2, LG343 *skn-1(zu135)/nT1[qIs51];Ex(ges-1-skn-1c–gfp)*, LG344 *geIs8[gpa-4-skn-1b–gfp]*, LG345 *geIs9[gpa-4-skn-1b–gfp]*, LG348 *skn-1(zu135)/ nT1[qIs51];geIs9*, LG349 *geIs10[ges-1-skn-1c–gfp]*, LG357 *skn-1(zu135)/nT1 [qIs51];geIs10*, CB1370 *daf-2(e1370)*, GR1307 *daf-16(mgDf50)*, CX3596 *kyIs128 [str-3-gfp]*, *LG347 daf-16(mgDf50);kyIs128*, CB1124 *che-3(e1124)*, PR813 *osm-5(p813)*, LG331 *lin-35(n745);Is007*, LG360 *skn-1(zu135)/nT1[qIs51];kyIs128*, FK181 *ksIs2[daf-7-gfp]*, LG359 *skn-1(zu135)/nT1[qIs51];ksIs2*, CX3594 *kyIs87 [srd-1-gfp]*, LG363 *skn-1(zu135)/nT1[qIs51];kyIs87*, CX3465 *kyIs39[sra-6-gfp]*, LG367 *skn-1(zu135)/nT1[qIs51];kyIs39*, LG313 *kyIs140[str-2-gfp]*, LG361 *skn-1(zu135)/nT1[qIs51];kyIs140*, AU10 *agIs1[dod-24-gfp]* (a gift of D. Kim and F. Ausubel), and the *Is007[skn-1–gfp]* and *Ex(gcs-1Δ2-gfp)* strains[10].

Transgenic strain construction. Extrachromosomal-array-carrying transgenic strains were generated using standard microinjection methods[32], injecting 50 ng μl^{-1} each of the transgene plasmid and the co-injection marker plasmid *pRF4*. Details of transgene construction are available on request.

The *gpa-4-skn-1b–gfp* transgene included 2.9 kb of *gpa-4* promoter fused to the full-length *skn-1b* complementary DNA and GFP. Adult animals carrying this transgene exhibited GFP fluorescence in the nuclei of the ASI neurons, and nowhere else. In larvae, weak expression was observed in one other pair of head neuron nuclei, which may be the AWAs[33]. Integrants of the *Ex(gpa-4-skn-1b–gfp)* transgenes, called *geIs8* and *geIs9*, were isolated independently from a standard γ-ray integration screen[34], and were backcrossed three times to N2 before analysis. Following integration, non-specific low-level GFP expression accumulated in the extreme posterior few gut nuclei of aged adult animals. It is improbable that this weak intestinal expression accounts for the observed effects on dietary restriction respiration in *geIs9* for several reasons: (1) the intestinal expression is very low-level, detectable only in aged adult animals on normal plates and not in our *ad libitum* or dietary restriction conditions, and limited to only a few posterior cells; (2) *skn-1b* is not normally present in the intestine (Fig. 2e, f) and lacks the transactivation domain of the *skn-1* isoforms that are normally present (Fig. 2a), and therefore is not necessarily capable of performing any normal intestinal function of *skn-1*; and (3) the low level of intestinal *skn-1b* that is present in *geIs9* animals is definitely not capable of rescuing the only intestine-specific function of *skn-1* that we are aware of, that is, arsenate resistance (Supplementary Fig. 6).

The *ges-1-skn-1c–gfp* transgene included 2.5 kb of *ges-1* promoter fused to the *skn-1c* cDNA and GFP. Twenty-one of 30 transgenic animals from the extrachromosomal transgenic line used in this report exhibited nuclear GFP fluorescence in the intestine following heat shock for 20 h at 29 °C (this degree of nuclear localization is comparable to that observed with the *Is007* integrant; Fig. 2e). No GFP fluorescence was ever observed outside the gut with or without heat shock. *geIs10*, an integrant of the *Ex(ges-1-skn-1c–gfp)* transgene was isolated from a standard γ-ray integration screen[34], and was backcrossed three times to N2 before analysis.

Dietary restriction protocol. Lifespan assays were performed in 6-well tissue culture plates, with each well containing 2.5 ml dietary restriction bottom medium and 2.5 ml dietary restriction top medium. Erythromycin that was added did not affect oxygen consumption rate or the time required for larval development (data not shown). The bacterial food source used was pPD129.36 (ref. 29). Bacterial concentration in a 1 ml sample of each well was monitored using a spectrophotometer and additional bacteria were added as necessary to maintain starting concentration. Worms were grown to the L4/young adult stage on NGM plates seeded with OP50 bacteria, then transferred individually into the wells and maintained at 20 °C with gentle gyratory shaking at 80 r.p.m.

For the progeny production assays, worms were cultured as described except that FuDR was excluded to allow progeny production, and worms were transferred daily to fresh wells.

Lifespan assay. Most lifespans in this report were measured in the liquid medium described above, except where noted; in these cases lifespans either were done on plates with the same composition as dietary restriction bottom medium seeded with 50 μl of concentrated bacteria ($\sim 5 \times 10^9$ c.f.u. ml^{-1}), or on standard NGM plates seeded with OP50 bacteria. Regardless of medium, the lifespan assay was performed by prodding individual animals with a worm pick every two days to determine when they died. Animals that were lost, or exploded, or died from internal hatching of progeny were censored at the time of the event[35]. In the assays shown in Fig. 4c–e, the dietary restriction medium contained 1% DMSO plus 0 μM, 30 μM or 60 μM myxothiazol, or 0.1% ethanol plus 0 μM or 1 μM antimycin.

Survival curve *P* values were calculated by the Mantel–Cox logrank test using Prism statistical software (Graphpad).

Laser ablation. The strain CX3596 *kyIs128[str-3-gfp]*, which expresses GFP specifically in the ASI neurons[27], was used to facilitate identification of the ASIs. Synchronized L1 animals were ablated, or mock-ablated in parallel, using a laser microbeam as described[26]. Worms were recovered to OP50 plates and grown to the L4/young adult stage, at which time absence of GFP was taken to indicate successful ablation. (We confirmed in a separate control experiment that a cohort of ablated animals lacking GFP also all lacked the ASI itself, and we never observed a GFP-negative animal in any of the hundreds of mock-ablated controls.) CX3596 proved to have a somewhat shorter lifespan than N2, but this was attributable to intrinsic strain differences and not to the mounting procedure (Supplementary Fig. 7).

Respiration assay. Mass OP50 plate cultures of synchronized worms were grown to the L4 stage, then transferred to Erlenmeyer flasks containing *ad libitum*-fed or dietary restriction bacterial concentrations. After three days of culture, animals were collected by centrifugation and washed free of bacteria in S buffer. Respiration rate was measured at 25 °C and normalized to protein content as described[28], using a Clark-type oxygen electrode (Microelectrodes Inc.). At least three 1 ml samples containing approximately 500 worms each were measured for each strain and condition in each experiment.

skn-1 mutants are maternal-effect lethal and must be maintained as balanced heterozygotes. To obtain nearly pure *skn-1* populations for the oxygen assays, we first balanced *skn-1(zu135)* with the reciprocal translocation *nT1[qIs51]*, which carries a transgene expressing GFP in the pharynx. We then isolated non-GFP-pharynx, *skn-1* homozygous animals from a synchronous L4 population using the COPAS Biosorter (Union Biometrica) before transfer to the *ad libitum* or dietary restriction culture conditions. This procedure yielded greater than 95% non-GFP-pharynx populations, which we confirmed to be *skn-1* homozygotes by verifying the embryonic lethality of progeny of samples of sorted worms.

RNAi feeding. RNAi feeding assays were done as described[29]. Worms were fed RNAi from L1 to L4. The *skn-1(RNAi)* construct has been described[36]. The *skn-1b(RNAi)* feeding vector included the first 209 nucleotides of the coding region of a *skn-1b* cDNA. The *skn-1c(RNAi)* feeding vector included the first 455 nucleotides of the coding region of a *skn-1c* cDNA.

Rapid amplification of cDNA ends. 5′ RACE (Invitrogen) was performed on total N2 RNA in accordance with the manufacturer's instructions.

Sudan Black fat staining. Day 3 adults cultured on plates, in *ad libitum* conditions, and in dietary restriction conditions were stained as described[37].

32. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10,** 3959–3970 (1991).
33. Kim, K., Colosimo, M. E., Yeung, H. & Sengupta, P. The UNC-3 Olf/EBF protein represses alternate neuronal programs to specify chemosensory neuron identity. *Dev. Biol.* **286,** 136–148 (2005).
34. Jin, Y. in *C. elegans: A Practical Approach* (ed. Hope, I. A.) 69–96 (Oxford University Press, New York, 1999).
35. Lawless, J. F. *Statistical Models and Methods for Lifetime Data* (Wiley, New York, 1982).
36. Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421,** 231–237 (2003).
37. McKay, R. M., McKay, J. P., Avery, L. & Graff, J. M. *C elegans*: a model for exploring the genetics of fat storage. *Dev. Cell* **4,** 131–142 (2003).

ARTICLES

PHA-4/Foxa mediates diet-restriction-induced longevity of *C. elegans*

Siler H. Panowski[1], Suzanne Wolff[1], Hugo Aguilaniu[1]†, Jenni Durieux[1] & Andrew Dillin[1]

Reduced food intake as a result of dietary restriction increases the lifespan of a wide variety of metazoans and delays the onset of multiple age-related pathologies. Dietary restriction elicits a genetically programmed response to nutrient availability that cannot be explained by a simple reduction in metabolism or slower growth of the organism. In the nematode worm *Caenorhabditis elegans,* the transcription factor PHA-4 has an essential role in the embryonic development of the foregut and is orthologous to genes encoding the mammalian family of Foxa transcription factors, *Foxa1*, *Foxa2* and *Foxa3*. Foxa family members have important roles during development, but also act later in life to regulate glucagon production and glucose homeostasis, particularly in response to fasting. Here we describe a newly discovered, adult-specific function for PHA-4 in the regulation of diet-restriction-mediated longevity in *C. elegans*. The role of PHA-4 in lifespan determination is specific for dietary restriction, because it is not required for the increased longevity caused by other genetic pathways that regulate ageing.

The insulin/IGF-1 signalling (IIS) pathway is a key regulator of the ageing process in worms, flies and mice, but its role in the regulation of diet-restriction-mediated longevity remains ambiguous[1–6]. Perfunctorily, it seems probable that the regulation of nutrient homeostasis and ageing by the IIS pathway might overlap with any regulatory networks affected by dietary restriction. However, prior research in worms suggests that diet-restriction-mediated increases in longevity can occur independently of the forkhead box O (FOXO) transcription factor DAF-16 (refs 7, 8), whereas the extended longevity of all known IIS mutants is completely dependent on DAF-16 (refs 4, 9–12); thus it seems unlikely that reduced food intake simply elicits an environment of reduced insulin signalling. From this hypothesis, we initially predicted that genetic components would not be shared across these two pathways.

Genetically, *smk-1* is an essential co-regulator of the longevity function of *daf-16*, and our previous data suggested a relationship in which these two genes cannot affect lifespan independently of each other[13]. *daf-16* is dispensable for the long lifespan of *eat-2(ad1116)* mutant animals (a genetic surrogate of dietary restriction exhibiting a reduced rate of pharyngeal pumping representative of eating)[8,14]. Thus, we were surprised to find that *smk-1* was required for the extended lifespan of *eat-2(ad1116)* mutant animals (see Methods, Fig. 1a and Supplementary Table 1). Consistent with previous results[7,8], we confirmed that *daf-16* is dispensable for diet-restriction-mediated longevity (Fig. 1b, Supplementary Fig. 1 and Supplementary Table 2). These data elicited the hypothesis that under conditions of low nutrient signalling, *smk-1* could interact genetically with a forkhead-like transcription factor other than *daf-16* to mediate the transcriptional response to dietary restriction. We systematically inactivated each of the fifteen forkhead-like genes found within the completed *C. elegans* genome[15] to examine their role in dietary restriction. RNA interference (RNAi) of only one, *pha-4*, completely suppressed the long lifespan of *eat-2(ad1116)* mutant animals (Fig. 1c and Supplementary Table 3). PHA-4 is orthologous to the human Foxa family of transcription factors (Supplementary Fig. 2)[16]. *Foxa1* homozygous mutant mice die shortly after birth, do not gain weight and are hypoglycaemic, suggesting an important role for *Foxa1* in pancreatic cell function and a central role in metabolic homeostasis[17,18]. *Foxa2* is also required for glucagon expression in the pancreas and induction of gluconeogenic genes during fasting in the liver[19]. *Foxa3* mutant mice become hypoglycaemic after a prolonged fasting[20,21]. The bifunctional role for Foxa family members in development and metabolic homeostasis of mammals prompted us to investigate further a potential role for *pha-4* in the regulation of metabolism and diet-restriction-mediated longevity of the adult worms in addition to its known role in development of the worm[22,23].

pha-4 is required for multiple forms of dietary restriction

Because *pha-4* is required for development of the worm pharynx, and *eat-2* mutations affect pharyngeal pumping rates, we tested whether a loss of *pha-4* suppressed dietary restriction in a non-genetic model. In the worm, dietary restriction can also be achieved by limiting the concentration of bacteria fed to worms in culture by bacterial dietary restriction (BDR)[24]. At high and extremely low food concentrations, wild-type animals are short lived, whereas conditions of optimal food intake result in increased longevity (see Methods, Fig. 1b and Supplementary Table 2). Much like wild-type animals, and in agreement with previous results[7], *daf-16(mu86)*-null mutant animals were longer lived at the optimal concentration and shorter lived at lower and higher concentrations, exhibiting a parabolic curve (Fig. 1b and Supplementary Fig. 1). In contrast, *pha-4(zu225);smg-1(cc546ts)* mutant worms, but not *smg-1(cc546)* control mutant worms (Supplementary Fig. 3), were short lived at all concentrations and did not exhibit a parabolic curve in response to varying food concentrations. A loss of *pha-4* fully blocked the entire response of lifespan to dietary restriction, as would be expected of a gene essential for diet-restriction-mediated longevity (Fig. 1b). In all experiments, transfer to restrictive temperatures to inactivate *pha-4* (ref. 22) as well as dietary restriction treatment itself was delayed until the first day of adulthood to avoid possible developmental abnormalities[22] (see

[1]The Salk Institute for Biological Studies, Molecular and Cell Biology Laboratory, 10010 North Torrey Pines Road, La Jolla, California 92037, USA.
†Present Address: Cellular and Organismal Aging Laboratory, Laboratoire de Biologie Moleculaire de la Cellule Ecole Normale Superieure de Lyon, 46 Allée d'Italie, 69364 Lyon, France.

Methods). We concurrently confirmed that a loss of *smk-1* using this method suppressed the extended lifespan under conditions of optimal dietary restriction (data not shown).

pha-4 is specific to diet-restriction-induced longevity

A loss of *pha-4* suppressed any potential lifespan extension across a spectrum of bacterial concentrations, suggesting that it was not causing a general sickness in the animal. However, to determine more conclusively whether *pha-4* was acting specifically to affect the dietary restriction pathway, we examined its effect on other pathways that influence longevity. Reduced IIS, by mutation of the insulin/IGF-1 receptor *daf-2* increases longevity. We found that *pha-4* was not required for the long lifespan of *daf-2* mutant animals. RNAi knockdown of *daf-16*, but not *pha-4*, completely suppressed the long lifespan of *daf-2(e1368)* (Fig. 1d), *daf-2(mu150)*, and *daf-2(e1370)* mutant animals (Supplementary Fig. 4a, b, respectively). Additionally, we tested whether *pha-4* was required for the long lifespan of animals with reduced mitochondrial electron transport chain activity[25,26]. Neither RNAi of *pha-4* nor the *pha-4(zu225);smg-1(cc546ts)* allele shortened the long lifespan of *cyc-1*-RNAi-treated animals (Fig. 1e and Supplementary Fig. 5a, respectively) or *isp-1(qm130)* mutant animals (Supplementary Fig. 5b) any more than reduction of *pha-4* in a wild-type background (Fig. 1f and Supplementary Table 1) and to a lesser extent than the loss of *daf-16* in *cyc-1*-RNAi-treated animals (Supplementary Fig. 5c). We thus conclude that *pha-4* is a specific requirement in the regulation of longevity in worms undergoing dietary restriction and that its loss does not simply cause a general sickness.

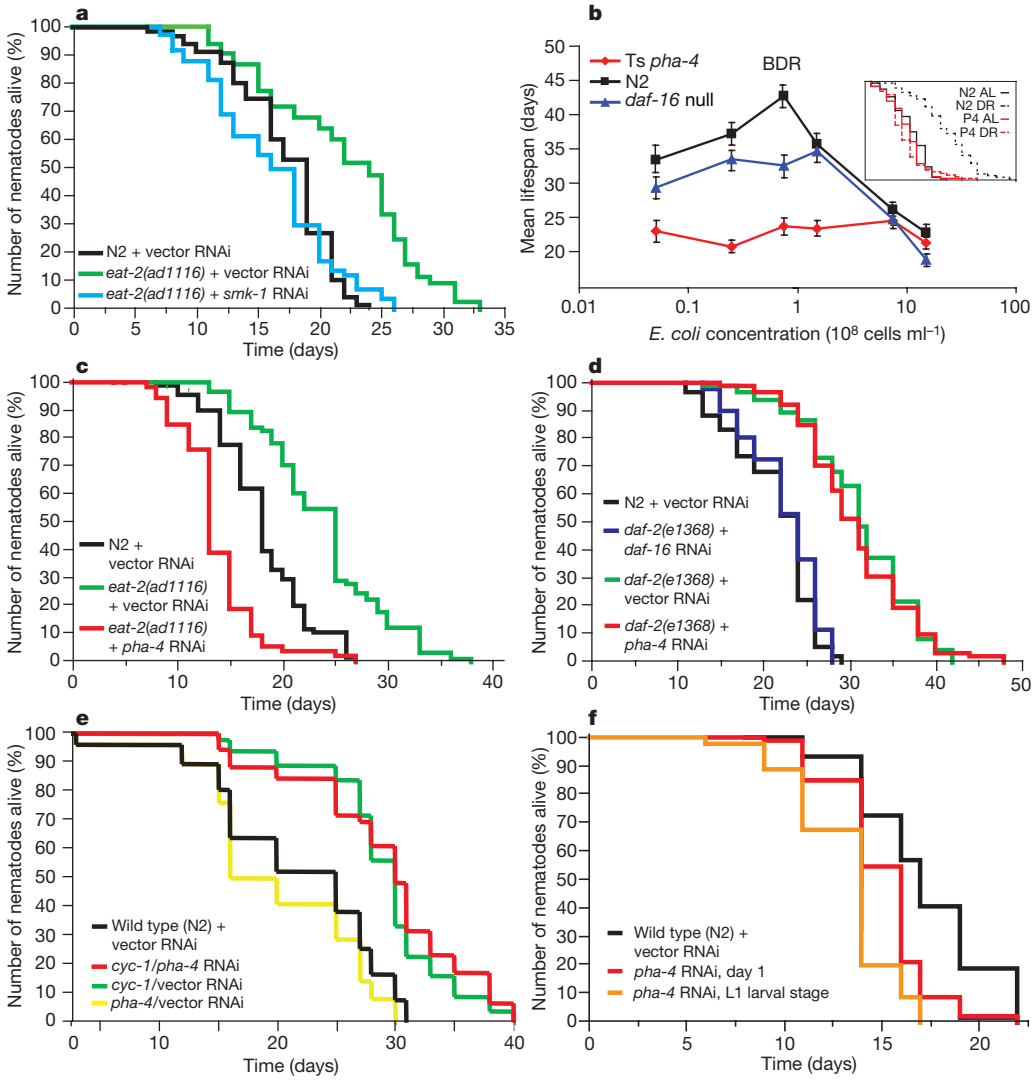


Figure 1 | *smk-1* and *pha-4* are required for diet-restriction-mediated longevity. All statistical data can be found in Supplementary Tables 1 and 2. Black lines indicate N2 worms grown on empty vector RNAi bacteria unless noted. **a**, *eat-2(ad1116)* worms fed empty vector RNAi bacteria (green line) lived significantly longer than *eat-2(ad1116)* worms fed *smk-1* RNAi bacteria (light blue line). **b**, Dietary restriction using non-RNAi bacterial dilution (BDR) results in a parabolic curve for wild-type worms (black line) and *daf-16(mu86)*-null mutant animals (blue line), but not *pha-4(zu225);smg-1(cc546ts)* temperature sensitive (ts) mutant animals (red line). Inset, lifespan plot of wild-type (N2; black line) and *pha-4(zu225);smg-1(cc546ts)* (P4; red line) worms with 7.5×10^8 cells ml^{-1} (*ad libitum*; AL) or 7.5×10^7 cells ml^{-1} (dietary restriction; DR). Error bars, s.e.m. **c**, *eat-2(ad1116)* mutant animals fed *pha-4* RNAi bacteria from the L1 larval stage (red line) were shorter lived than animals fed vector RNAi bacteria (green line). **d**, *daf-2(e1368)* mutant animals fed either vector RNAi bacteria (green line) or *pha-4* RNAi bacteria (red line) lived significantly longer than when fed *daf-16* RNAi bacteria (blue line). **e**, Wild-type (N2) animals fed 50% *cyc-1* and 50% vector RNAi bacteria (green line) or 50% *cyc-1* and 50% *pha-4* RNAi bacteria (red line) showed a similar lifespan extension compared to N2 animals fed vector RNAi bacteria alone (black line). N2 animals fed 50% *pha-4* and 50% vector RNAi bacteria (yellow line) had a slightly shorter lifespan. Lifespan analyses of animals with reduced mitochondrial electron transport chain were performed at 15 °C. **f**, Wild-type worms fed *pha-4* RNAi bacteria starting from either day 1 of adulthood (red line) or the L1 larval stage (orange line) lived significantly shorter than worms fed vector RNAi bacteria (black line).

The role of *pha-4* in development and longevity is separable

pha-4 has an essential early role during embryo development in the morphogenesis of the pharynx, and inactivation of *pha-4* up to the first larval stage, L1, can result in lethality[23]. We tested whether the early developmental function of *pha-4* can be temporally separated from its role in dietary restriction during adulthood. We allowed *eat-2(ad1116)* mutant animals to develop through the larval stages (L1–L4) and grow on normal bacteria and then shifted the animals on the first day of adulthood to bacteria expressing *pha-4* double-stranded-RNA, thereby only inactivating *pha-4* during adulthood—long after pharyngeal development had completed[23]. RNAi of *pha-4* during only adulthood suppressed the long lifespan of *eat-2(ad1116)* mutant animals to wild-type levels (Fig. 2). In support of these data, *pha-4(zu225);smg-1(cc546ts)* mutant worms in our BDR experiments were not shifted to the restrictive temperature to inactivate *pha-4* (ref. 22) until adulthood.

We additionally considered whether reduction of *pha-4* could suppress diet-restriction-mediated longevity by altering pharyngeal function during adulthood by indirectly affecting the feeding rates of animals and pushing diet-restriction animals towards starvation. We found this hypothesis inconsistent with multiple observations. First, the pumping (feeding) rate of wild-type animals grown on control bacteria was very similar to the pumping rate of wild-type animals treated with *pha-4* RNAi (wild type treated with vector RNAi, 242 ± 10 pumps per min (± s.d.); wild type treated with *pha-4* RNAi, 238.2 ± 11.5 pumps per min (± s.d.)). Additionally, *eat-2(ad1116)* mutant animals treated with *pha-4* RNAi did not exhibit altered feeding rates (*eat-2(ad1116)* treated with vector RNAi, 50.1 ± 7.1 pumps per min; *eat-2(ad1116)* treated with *pha-4* RNAi, 48.4 ± 5.5 pumps per min). We confirmed that RNAi of *pha-4* was activated by the time pumping rates were monitored by following the green fluorescent protein (GFP) signal of *pha-4–gfp* transgenic animals treated with *pha-4* RNAi (see Methods and Supplementary Fig. 6). In agreement with this observation, *pha-4* RNAi did not increase longevity of wild-type animals, as would be expected if feeding rates were reduced[8]. In fact, RNAi of *pha-4* slightly shortened wild-type longevity, even when applied specifically to adult animals (Fig. 1f). As noted previously, the *pha-4(zu225)* mutation did not change the parabolic relationship observed between BDR and longevity: it blocked the entire response (Fig. 1b). Finally, *pha-4* RNAi did not further increase the long lifespan of *daf-2* mutant animals (Fig. 1d and Supplementary Fig. 4a, b), as is observed with *eat-2;daf-2* mutant animals that live longer than either single mutation[8], and did not enhance the long lifespan of animals treated with RNAi or mutation resulting in reduced electron transport chains (Fig. 1e and Supplementary Fig. 5a, b).

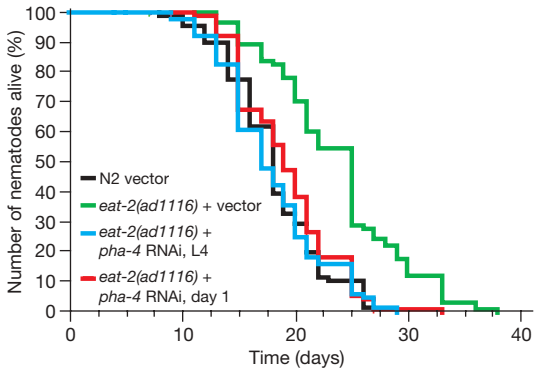


Figure 2 | *pha-4* is required during adulthood to regulate longevity in response to dietary restriction. *eat-2(ad1116)* mutants were transferred to *pha-4* RNAi bacteria at the L4 larval stage (blue line; mean lifespan 18 ± 0.5 days; mean ± s.e.m.) or day 1 of adulthood (red line; mean lifespan 19.2 ± 0.4 days) and in both cases showed a decreased lifespan compared with *eat-2(ad1116)* animals fed vector RNAi bacteria (green line; mean lifespan 23.8 ± 0.6 days). Mean lifespan of N2 worms fed vector RNAi bacteria (black line) was 18.1 ± 0.4 days.

pha-4 expression is increased in response to dietary restriction

pha-4 is expressed in the developing pharynx and intestine during embryogenesis and larval stages[23,27]. We asked whether the expression pattern of *pha-4* during adulthood was different from its developmental expression pattern. Using a red fluorescent protein (RFP) transcriptional fusion to the *pha-4* promoter, we observed strong expression in the developing pharynx and in the intestine, as noted previously[23,27]. In the adult animal, expression was lacking in the pharynx, but still present in the intestine (Fig. 3a). Using a full-length *pha-4* complementary DNA translation fusion to GFP under the *pha-4* promoter, we observed nuclear localization of PHA-4 during development and adulthood within the same cells (Fig. 3b and see Methods) and also found *pha-4* expression in the adult worm expanded to a few neuronal cells in the head and tail, which were not found in the developing animal (Fig. 3b). This expression pattern did not change in response to dietary restriction (data not shown), and PHA-4 seemed constitutively nuclear under all conditions tested (Fig. 3c).

During embryogenesis, levels of PHA-4 expression determine its binding specificity: low levels of PHA-4 bind high-affinity sites in promoters during early embryogenesis; PHA-4 does not bind to low-affinity sites until late in embryogenesis when *pha-4* expression levels increase[22]. Following this paradigm, we reasoned that expression of *pha-4* might increase during dietary restriction to facilitate its binding to diet-restriction-specific genes. Using both semi-quantitative PCR with reverse transcription (RT–PCR) and quantitative real-time RT–PCR (Q-PCR), expression of *pha-4* increased by more than 80% in response to dietary restriction (Fig. 3d).

Overexpression of *pha-4* extends longevity in the absence of *daf-16*

Because expression levels of *pha-4* were increased in response to dietary restriction, we tested whether overexpression of *pha-4* was sufficient to extend longevity under normal feeding conditions. Eleven independent lines overexpressing *pha-4* were established (see Methods). In nine lines, *pha-4* overexpression increased longevity of wild-type animals, but only slightly (Supplementary Table 4). However, when the same *pha-4* expression construct was used to overexpress *pha-4* in a *daf-16(mu86)*-null mutant strain, we observed a statistically significant increase in lifespan (Fig. 4 and Supplementary Table 4). There seem to be at least two explanations for this result. One, an inherent competition between *daf-16* and *pha-4* in wild-type animals may exist; or two, the role of *daf-16* and *pha-4* may be partially redundant in determination of longevity of wild-type animals. In any event, the relative increase in lifespan by *pha-4* overexpression was greatest in the complete absence of *daf-16*.

The *sod* gene family is differentially regulated by DAF-16 and PHA-4

In analysing the potential competition among *daf-16* and *pha-4*, we noticed that the consensus DNA binding sites for DAF-16 and PHA-4 overlap: PHA-4, T(A/G)TT(T/G)(A/G)(T/C) (ref. 22) versus DAF-16, T(A/G)TTTAC (ref. 28). This observation raised the hypothesis that DAF-16 and PHA-4 regulate expression of the same genes either directly or indirectly. *sod-3*, a mitochondrial Fe/Mn superoxide dismutase[29–31], is the best characterized DAF-16 target gene and contains three DAF-16 DNA binding sites within the promoter region[32]. All three DAF-16 sites overlap with the consensus PHA-4 DNA binding site. Therefore, we used Q-PCR analysis to examine *sod-3* expression in diet-restricted, *eat-2(ad1116)* mutant animals. Surprisingly, we found no increase in expression levels of *sod-3* in response to dietary restriction (Fig. 5a, see Methods). Furthermore, the basal level of *sod-3* expression was not altered in diet-restricted animals lacking *pha-4* (Fig. 5b).

The *sod-1* promoter contains four consensus PHA-4 binding sites. Furthermore, the mouse *sod-1* orthologue, *sod-1*, has been shown to be a transcriptional target of *Foxa1* (ref. 33). *sod-1* is a cytoplasmic Cu/Zn superoxide dismutase[34]. We tested whether the *C. elegans*

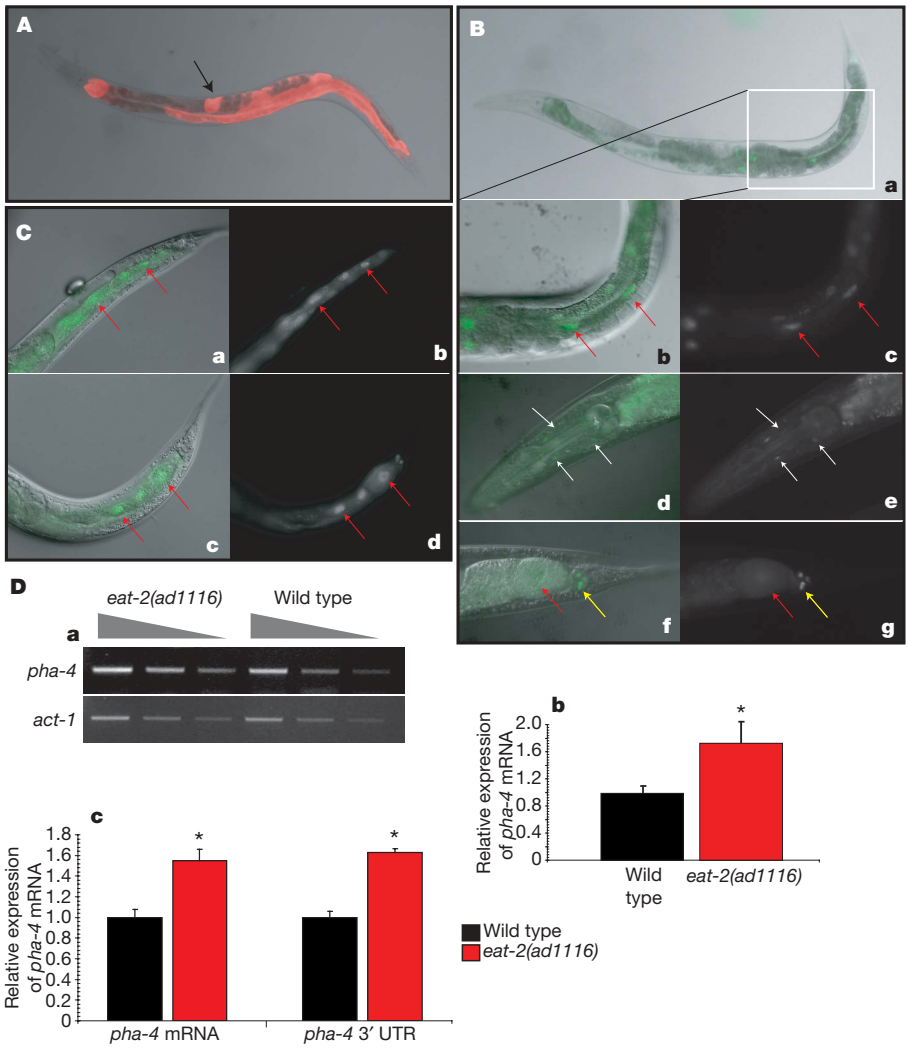


Figure 3 | Regulation and localization of *pha-4* in response to dietary restriction. **A**, A *pha-4–rfp* transcriptional fusion construct (see Methods) reveals *pha-4* expression in the intestine and somatic gonad (spermatheca, black arrow) of the adult worm. **B**, **a**, AD84 worms reveal PHA-4–GFP nuclear localization in intestinal cells (**b** and **c**, red arrows), head neurons (**d** and **e**, white arrows) and tail neurons (**f** and **g**, yellow arrows). **C**, AD84 worms carrying the *pha-4–gfp* transgene were placed in high-food cultures (7.5×10^8 cells ml^{-1}; **a**, **b**) or low-food cultures (7.5×10^7 cells ml^{-1}; **c**, **d**). Nuclear localization of PHA-4–GFP in intestinal nuclei (red arrows) remained constant under both conditions (see Methods). Images were taken on day four of adulthood. **D**, Semi-quantitative RT–PCR (**a**) reveals an increase in *pha-4* messenger RNA levels in *eat-2(ad1116)* animals compared with wild-type animals of ~80% (unpaired two-tailed *t*-test *P*-value < 0.001; error bars, s.d.; **b**). Q-PCR analysis of the *pha-4* gene and the *pha-4* 3′ UTR confirms that *pha-4* mRNA levels were increased in *eat-2(ad1116)* worms (red bars) compared with wild-type worms (black bars) (unpaired two-tailed *t*-test *P*-value < 0.0001; errors bars, s.d.; **c**). $n = 3$ for both semi-quantitative RT–PCR and Q-PCR experiments.

sod-1 was transcriptionally regulated in *eat-2(ad1116)* mutant animals. By Q-PCR analysis, *sod-1* expression was greatly upregulated in response to dietary restriction (Fig. 5c). In diet-restricted animals, *sod-1* expression was decreased in the absence of *pha-4*, but was slightly increased in the absence of *daf-16* (Fig. 5d). Therefore, expression of *sod-1* requires PHA-4, but not DAF-16, in *eat-2(ad1116)* mutant animals.

The *C. elegans* genome contains five *sod* genes including *sod-3* and *sod-1* (ref. 15). We further investigated the expression patterns of each of the *sod* gene family members under conditions of dietary restriction or reduced IIS signalling. Interestingly, we found that the expression level of every *sod* gene except for *sod-3* was increased under dietary restriction (Fig. 5e). The increases were *pha-4*-dependent (Fig. 5f). In response to reduced IIS, *sod-1*, *sod-3* and *sod-5* expression levels were increased (Fig. 5g); this increase was *daf-16*-dependent (Fig. 5h). Taken together, we thus find that *sod-2* and sod-*4* expression is specific to dietary restriction and dependent on *pha-4*, whereas *sod-3* expression is specific to reduced IIS and dependent on *daf-16*. Common to both dietary restriction and reduced IIS, expression of *sod-1* and *sod-5* are increased by PHA-4 and DAF-16, respectively (Fig. 5i). Although each *sod* gene contains respective predicted DAF-16 and PHA-4 binding sites within their promoters, regulation by additional factors cannot be ruled out at this time.

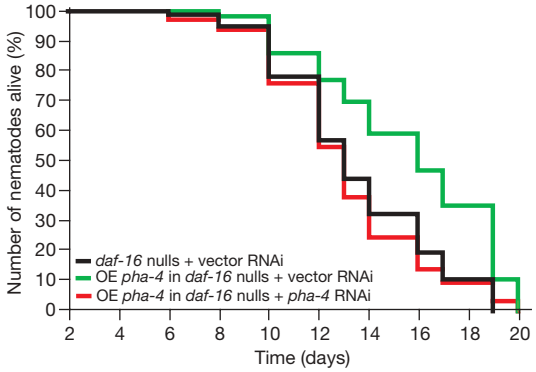


Figure 4 | Increased dosage of *pha-4* extends lifespan. Transgenic *daf-16(mu86)*-null mutant worms carrying an overexpressor (OE) *pha-4* transgene (AD115, green line) were long-lived compared with *daf-16(mu86)* worms (AD105, black line). This lifespan extension was fully suppressed by *pha-4* RNAi (red line). Statistical data can be found in Supplementary Table 4.

Discussion

In worms, PHA-4 is bifunctional, having an early developmental function in pharyngeal determination during embryogenesis and the L1 larval stage[22,23], and a later function during adulthood in regulating the response to dietary restriction. This dual mode of action of PHA-4 is similar to that of DAF-16, which is required during early larval stages to regulate the dauer developmental decision and reproductive status of the animal, and later during adulthood to regulate the response of ageing to IIS[35]. In mammals, a parallel regulation of insulin levels by FOXO proteins[36], and glucagon levels by *Foxa1* and *Foxa2* (refs 18, 19), supports a model in which, under continually low nutrient signalling, PHA-4/Foxa may mediate levels of glucagon or other changes in hormones ultimately capable of regulating the ageing process. In contrast,

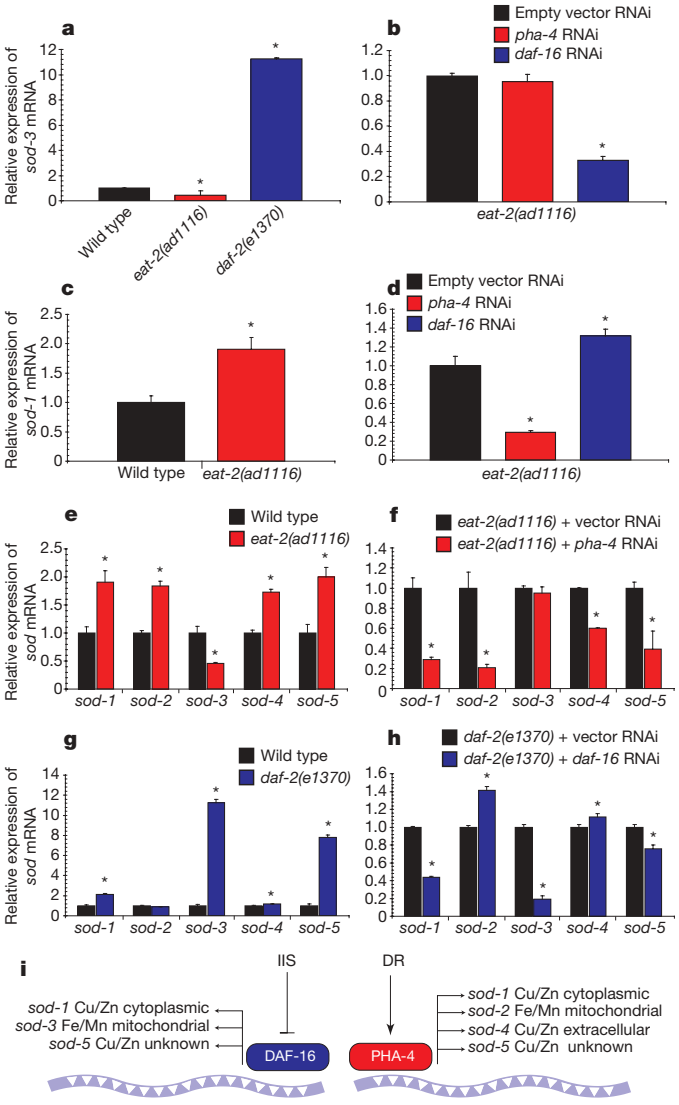


Figure 5 | Differential transcriptional regulation of *sod*s by *pha-4* and *daf-16* in response to dietary restriction and IIS. All mRNA expression levels were determined using Q-PCR analysis and performed in parallel (see Methods). Q-PCR reactions were run in quadruplicate and averages from one representative set of reactions are depicted in graphs. Error bars represent s.d. for the reaction depicted and asterisks indicate a change in expression with an unpaired two-tailed *t*-test *P*-value <0.005 as compared with black bars of the same graph and gene. **a**, **c**, **e**, **g**, mRNA levels of worms fed vector RNAi bacteria. **a**, *sod-3* expression levels were increased in *daf-2(e1370)* animals (blue bar) compared with wild-type N2 (black bar) and *eat-2(ad1116)* animals (red bar). **b**, *sod-3* mRNA expression levels in *eat-2(ad1116)* mutant worms were unaffected by *pha-4* RNAi (red bar), but were decreased in response to *daf-16* RNAi (blue bar). **c**, *sod-1* mRNA levels were increased in *eat-2(ad1116)* worms (red bar) compared with wild-type worms (black bar). **d**, *sod-1* mRNA expression levels were decreased in *eat-2(ad1116)* worms fed *pha-4* RNAi (red bar). **e**, All *sod*s, except *sod-3*, were upregulated in *eat-2(ad1116)* mutant animals (red bars) compared with wild-type animals (black bars). **f**, *sod-1*, *sod-2*, *sod-4* and *sod-5* mRNA expression in *eat-2(ad1116)* worms was greatly decreased when worms were fed *pha-4* RNAi bacteria (red bars) compared with *eat-2(ad1116)* worms fed vector RNAi bacteria (black bars). **g**, *daf-2(e1370)* mutant worms (blue bars) had elevated levels of *sod-1*, *sod-3* and *sod-5* mRNA when compared with wild-type worms (black bars). **h**, *daf-2(e1370)* mutant animals fed *daf-16* RNAi bacteria (blue bars) had reduced levels of *sod-1*, *sod-3* and *sod-5* compared with *daf-2(e1370)* mutant animals fed vector RNAi (black bars). **i**, Model depicting differential regulation of *sod*s in response to IIS and dietary restriction (DR) mediated by DAF-16 and PHA-4, respectively. Types (Fe/Mn, Cu/Zn) and location, if known, of superoxide dismutases are listed next to the genes[30,31,45,46].

in times of severe stress or starvation, DAF-16/FOXO will mediate the response to decreased insulin signalling. Although *C. elegans* does not contain an obvious glucagon orthologue, it does contain a full complement of insulin-like peptides[37], suggesting that a conserved functional regulation of glucose homeostasis may be present. The finding that some insulin-like peptides work as agonists[38], whereas others are antagonists[39], to insulin signalling in worms indicates that glucose homeostasis could be more directly regulated by expression of insulin-like peptides in response to dietary restriction. In the future, it will be imperative to understand whether the Foxa family is required for diet-restriction-mediated longevity in mammals and what part glucagon production plays in this process.

The response to IIS involves the DAF-16-dependent regulation of *sod-1, sod-3* and *sod-5*, whereas dietary restriction involves the PHA-4-dependent expression of *sod-1, sod-2, sod-4* and *sod-5*. The disparate transcriptional outcomes of these treatments on oxygen radical scavenging genes could suggest that a different form of reactive oxygen species production may be induced under conditions of reduced IIS than is induced under conditions of dietary restriction. This may indicate divergent underlying metabolic consequences stemming from the manipulation of these independent pathways. Alternatively, as the expression patterns for most of these *sod*s remain unknown, the differential transcriptional regulation of *sod*s under IIS and dietary restriction could indicate distinct tissue-specific requirements for IIS and diet-restriction-mediated longevity. In *C. elegans*, IIS is required in the neurons and intestinal cells to regulate lifespan[40–42]. Although expression patterns of *pha-4* overlap with those of *daf-16* in the intestine and some neuronal cells, it is not known which tissues integrate and respond to reduced dietary intake. It is possible that the same tissues that exhibit increased levels of oxidative-stress response genes also will require DAF-16 or PHA-4 to affect longevity. It is likely that *sod* gene regulation is not the sole target of DAF-16 and PHA-4 for longevity assurance, but rather these transcription factors orchestrate a larger regulatory network that has been previously proposed[38,43].

Many of the physiological outcomes of animals with reduced IIS compared with animals undergoing dietary restriction are similar, including reduced body size, lower plasma IGF-1 and insulin levels, and increased insulin sensitivity. Furthermore, transcriptional profiling of long-lived dwarf mice, having reduced IGF-1 signalling, in combination with dietary restriction, additively increased expression of multiple liver-specific genes[44]. However, compelling genetic analysis indicates that many key differences among IIS- and diet-restricted mice exist as well. For example, long-lived growth-hormone-deficient mice can still respond to dietary restriction, and IGF-1R long-lived heterozygous mice do not show protracted or reduced reproduction[3].

Therefore, given the discrepancy between the mode of action elicited by reduced IIS and dietary restriction that results in increased longevity of an organism, it is important to note that *pha-4* is exceptionally specific for the longevity induced by dietary restriction. Reduction of *pha-4* does not suppress the long lifespan of *daf-2* mutant animals or animals with defective electron transport chains. We conclude, in agreement with previous reports[7,8], that there exists an independent pathway for the regulation of dietary restriction in worms. Consistent with this observation, worms undergoing dietary restriction do not require *daf-16*. Our results instead suggest that dietary restriction impinges on an independent mechanism that ultimately increases the activity of PHA-4. Overexpression of *pha-4* extends longevity in the absence of *daf-16*, and *pha-4* expression is increased under conditions of dietary restriction. We thus report the first findings of a forkhead transcription factor that acts independent of and with a parallel mechanism to *daf-16* and IIS to regulate the ageing process in diet-restricted worms.

METHODS SUMMARY

A detailed description of all experimental methods is provided in the Methods section.

C. elegans strains, growth, imaging, lifespan analysis, Q-PCR and RNAi application were performed as previously described[13]. For bacterial restriction

studies, each lifespan consisted of 4 wells, with 1 ml of culture and 15 worms per well ($n = 60$). Lifespans were scored and worms transferred to new cultures every 3–4 days. Liquid cultures were prepared using an overnight culture of OP50 *Escherichia coli* grown at 37 °C. Bacteria were washed three times in S-Basal medium. The bacterial concentration was adjusted to 1.5×10^{-9} cells ml^{-1} in S-Basal medium containing cholesterol, carbenicillin, tetracycline and kanamycin (for concentrations see Methods). Serial dilutions were performed to achieve bacterial concentrations of 7.5×10^{8}, 1.5×10^{8}, 7.5×10^{7}, 2.5×10^{7} and 5×10^{6} cells ml^{-1}. Cultures contained Fluorodeoxyuridine (FUDR) at 100 μg ml^{-1} for the first twelve days of lifespan analysis to block worm reproduction. For analysis of the temperature-sensitive *pha-4(zu225)* mutant allele, *pha-4(zu225);smg-1(cc546ts)*[22] double mutant worms were grown at 25 °C to inactivate *smg-1* and allow production of functional *pha-4*. *pha-4* was inactivated by shifting double mutants to 15 °C, restoring *smg-1* activity, which results in degradation of the *pha-4(zu225)* allele after the first day of adulthood, thus avoiding any developmental defects owing to loss of *pha-4* during larval stages. All control worms were treated identically.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 September 2006; accepted 12 April 2007.
Published online 2 May 2007.

1. Clancy, D. J., Gems, D., Hafen, E., Leevers, S. J. & Partridge, L. Dietary restriction in long-lived dwarf flies. *Science* **296,** 319 (2002).
2. Clancy, D. J. *et al.* Extension of life-span by loss of CHICO, a *Drosophila* insulin receptor substrate protein. *Science* **292,** 104–106 (2001).
3. Holzenberger, M. *et al.* IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. *Nature* **421,** 182–187 (2003).
4. Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. A. C. *elegans* mutant that lives twice as long as wild type. *Nature* **366,** 461–464 (1993).
5. Tatar, M. *et al.* A mutant *Drosophila* insulin receptor homolog that extends life-span and impairs neuroendocrine function. *Science* **292,** 107–110 (2001).
6. Tu, M. P., Epstein, D. & Tatar, M. The demography of slow aging in male and female *Drosophila* mutant for the insulin-receptor substrate homologue chico. *Aging Cell* **1,** 75–80 (2002).
7. Houthoofd, K., Braeckman, B. P., Johnson, T. E. & Vanfleteren, J. R. Life extension via dietary restriction is independent of the Ins/IGF-1 signalling pathway in *Caenorhabditis elegans*. *Exp. Gerontol.* **38,** 947–954 (2003).
8. Lakowski, B. & Hekimi, S. The genetics of caloric restriction in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **95,** 13091–13096 (1998).
9. Henderson, S. T. & Johnson, T. E. *daf-16* integrates developmental and environmental inputs to mediate aging in the nematode *Caenorhabditis elegans*. *Curr. Biol.* **11,** 1975–1980 (2001).
10. Lin, K., Dorman, J. B., Rodan, A. & Kenyon, C. *daf-16*: An HNF-3/forkhead family member that can function to double the life-span of *Caenorhabditis elegans*. *Science* **278,** 1319–1322 (1997).
11. Lin, K., Hsin, H., Libina, N. & Kenyon, C. Regulation of the *Caenorhabditis elegans* longevity protein DAF-16 by insulin/IGF-1 and germline signaling. *Nature Genet.* **28,** 139–145 (2001).
12. Ogg, S. *et al.* The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. *Nature* **389,** 994–999 (1997).
13. Wolff, S. *et al.* SMK-1, an essential regulator of DAF-16-mediated longevity. *Cell* **124,** 1039–1053 (2006).
14. Avery, L. The genetics of feeding in *Caenorhabditis elegans*. *Genetics* **133,** 897–917 (1993).
15. *C. elegans* sequencing consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282,** 2012–2018 (1998).
16. Horner, M. A. *et al. pha-4*, an *HNF-3* homolog, specifies pharyngeal organ identity in *Caenorhabditis elegans*. *Genes Dev.* **12,** 1947–1952 (1998).
17. Shih, D. Q., Navas, M. A., Kuwajima, S., Duncan, S. A. & Stoffel, M. Impaired glucose homeostasis and neonatal mortality in hepatocyte nuclear factor 3α-deficient mice. *Proc. Natl Acad. Sci. USA* **96,** 10152–10157 (1999).
18. Kaestner, K. H., Katz, J., Liu, Y., Drucker, D. J. & Schutz, G. Inactivation of the winged helix transcription factor HNF3α affects glucose homeostasis and islet glucagon gene expression *in vivo*. *Genes Dev.* **13,** 495–504 (1999).
19. Zhang, L., Rubins, N. E., Ahima, R. S., Greenbaum, L. E. & Kaestner, K. H. Foxa2 integrates the transcriptional response of the hepatocyte to fasting. *Cell Metab.* **2,** 141–148 (2005).
20. Kaestner, K. H., Hiemisch, H. & Schutz, G. Targeted disruption of the gene encoding hepatocyte nuclear factor 3γ results in reduced transcription of hepatocyte-specific genes. *Mol. Cell. Biol.* **18,** 4245–4251 (1998).
21. Shen, W., Scearce, L. M., Brestelli, J. E., Sund, N. J. & Kaestner, K. H. *Foxa3* (hepatocyte nuclear factor 3γ) is required for the regulation of hepatic GLUT2 expression and the maintenance of glucose homeostasis during a prolonged fast. *J. Biol. Chem.* **276,** 42812–42817 (2001).
22. Gaudet, J. & Mango, S. E. Regulation of organogenesis by the *Caenorhabditis elegans* FoxA protein PHA-4. *Science* **295,** 821–825 (2002).
23. Mango, S. E., Lambie, E. J. & Kimble, J. The *pha-4* gene is required to generate the pharyngeal primordium of *Caenorhabditis elegans*. *Development* **120,** 3019–3031 (1994).
24. Klass, M. R. Aging in the nematode *Caenorhabditis elegans*: major biological and environmental factors influencing life span. *Mech. Ageing Dev.* **6,** 413–429 (1977).
25. Dillin, A. *et al.* Rates of behavior and aging specified by mitochondrial function during development. *Science* **298,** 2398–2401 (2002).
26. Lee, S. S. *et al.* A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity. *Nature Genet.* **33,** 40–48 (2003).
27. Azzaria, M., Goszczynski, B., Chung, M. A., Kalb, J. M. & McGhee, J. D. A fork head/HNF-3 homolog expressed in the pharynx and intestine of the *Caenorhabditis elegans* embryo. *Dev. Biol.* **178,** 289–303 (1996).
28. Furuyama, T., Nakazawa, T., Nakano, I. & Mori, N. Identification of the differential distribution patterns of mRNAs and consensus binding sequences for mouse DAF-16 homologues. *Biochem. J.* **349,** 629–634 (2000).
29. Giglio, M. P., Hunter, T., Bannister, J. V., Bannister, W. H. & Hunter, G. J. The manganese superoxide dismutase gene of *Caenorhabditis elegans*. *Biochem. Mol. Biol. Int.* **33,** 37–40 (1994).
30. Hunter, T., Bannister, W. H. & Hunter, G. J. Cloning, expression, and characterization of two manganese superoxide dismutases from *Caenorhabditis elegans*. *J. Biol. Chem.* **272,** 28652–28659 (1997).
31. Suzuki, N., Inokuma, K., Yasuda, K. & Ishii, N. Cloning, sequencing and mapping of a manganese superoxide dismutase gene of the nematode *Caenorhabditis elegans*. *DNA Res.* **3,** 171–174 (1996).
32. Honda, Y. & Honda, S. The *daf-2* gene network for longevity regulates oxidative stress resistance and Mn-superoxide dismutase gene expression in *Caenorhabditis elegans*. *FASEB J.* **13,** 1385–1393 (1999).
33. Carroll, J. S. *et al.* Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122,** 33–43 (2005).
34. Giglio, A. M., Hunter, T., Bannister, J. V., Bannister, W. H. & Hunter, G. J. The copper/zinc superoxide dismutase gene of *Caenorhabditis elegans*. *Biochem. Mol. Biol. Int.* **33,** 41–44 (1994).
35. Dillin, A., Crawford, D. K. & Kenyon, C. Timing requirements for insulin/IGF-1 signaling in *C. elegans*. *Science* **298,** 830–834 (2002).
36. Puig, O. & Tjian, R. Transcriptional feedback control of insulin receptor by dFOXO/FOXO1. *Genes Dev.* **19,** 2435–2446 (2005).
37. Pierce, S. B. *et al.* Regulation of DAF-2 receptor signaling by human insulin and *ins-1*, a member of the unusually large and diverse *C. elegans* insulin gene family. *Genes Dev.* **15,** 672–686 (2001).
38. Murphy, C. T. *et al.* Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature* **424,** 277–283 (2003).
39. Li, W., Kennedy, S. G. & Ruvkun, G. *daf-28* encodes a *C. elegans* insulin superfamily member that is regulated by environmental cues and acts in the DAF-2 signaling pathway. *Genes Dev.* **17,** 844–858 (2003).
40. Apfeld, J. & Kenyon, C. Regulation of lifespan by sensory perception in *Caenorhabditis elegans*. *Nature* **402,** 804–809 (1999).
41. Libina, N., Berman, J. R. & Kenyon, C. Tissue-specific activities of *C. elegans* DAF-16 in the regulation of lifespan. *Cell* **115,** 489–502 (2003).
42. Wolkow, C. A., Kimura, K. D., Lee, M. S. & Ruvkun, G. Regulation of *C. elegans* life-span by insulinlike signaling in the nervous system. *Science* **290,** 147–150 (2000).
43. McElwee, J., Bubb, K. & Thomas, J. H. Transcriptional outputs of the *Caenorhabditis elegans* forkhead protein DAF-16. *Aging Cell* **2,** 111–121 (2003).
44. Tsuchiya, T. *et al.* Additive regulation of hepatic gene expression by dwarfism and caloric restriction. *Physiol. Genomics* **17,** 307–315 (2004).
45. Fujii, M., Ishii, N., Joguchi, A., Yasuda, K. & Ayusawa, D. A novel superoxide dismutase gene encoding membrane-bound and extracellular isoforms by alternative splicing in *Caenorhabditis elegans*. *DNA Res.* **5,** 25–30 (1998).
46. Larsen, P. L. Aging and resistance to oxidative damage in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **90,** 8905–8909 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Dillin laboratory for critical comments on this work; M. Montminy and R. Shaw for comments on the manuscript; and S. Mango for reagents and discussion during the course of this work. We thank R. Tsien for use of dTOMATO. This work was supported by grants from the NIH/NIA, The Ellison Medical Foundation and the American Diabetes Association. H.A. thanks the Jane Coffin Childs Foundation for support. A.D. is founder of Proteoguard Pharmaceuticals.

Author Contributions A.D., S.W., H.A. and S.P. conceived the framework of the manuscript. A.D., S.W. and S.P. wrote the paper. A.D. oversaw the entire project. S.W. and H.A. screened all forkhead-related genes for their role in dietary restriction. J.D. performed the *isp-1(qm150)* experiments. S.P. created all transgenic lines, performed overexpression, localization, BDR and Q-PCR experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.D. (dillin@salk.edu).

METHODS

***C. elegans* methods and generation of transgenic lines.** CF1037 (*daf-16(mu86)I*), CF1041 (*daf-2(e1370)III*), DR1572 (*daf-2(e1368)III*), CF512 (*fer-15(b26)II;fem-1(hc17)IV*), CF1379 (*daf-2(mu150)*), PD8120 (*smg-1(cc546ts)I*) and wild-type *C. elegans* (N2) strains were obtained from the *Caenorhabditis* Genetic Center. SM190 (*pha-4(zu225);smg-1(cc546ts)*)[22] was provided by S. Mango. Nematodes were maintained and handled using standard methods[47]. For generation of transgenic animals, plasmid DNA containing the construct of interest and pRF4(*rol-6*)[48] was microinjected into the gonads of adult hermaphrodite animals, using standard methods[48]. F_1 progeny were selected on the basis of the roller phenotype. Individual F_2 worms were isolated to establish independent lines.

For generation of worm strains AD137–AD141 and AD143–AD148 (N2, *pha-4*; N2 worms overexpressing a *pha-4* transgene), the plasmid DNA mix consisted of 50 ng μl^{-1} pSP15(*pha-4*) and 50 ng μl^{-1} pRF4(*rol-6*). Wild-type (N2) worms were injected. Worms used as controls in lifespan experiments for AD137–AD141 and AD143–AD148 were generated by microinjecting wild-type (N2) worms with 50 ng μl^{-1} of pRF4(*rol-6*) alone.

For generation of AD115 (*daf-16(mu86);pha-4*), the plasmid mix consisted of 50 ng μl^{-1} pSP15*(pha-4)* and 50 ng μl^{-1} pRF4(*rol-6*). *daf-16(mu86)* worms were injected. *daf-16(mu86)* animals injected with 50 ng μl^{-1} of pRF4(*rol-6*) alone were used as a control in lifespan experiments and termed AD105.

For AD150 (N2, *pha-4-rfp*), wild-type (N2) worms were microinjected with plasmid DNA containing 75 ng μl^{-1} pSP23(*rfp*) and 75 ng μl^{-1} pRF4(*rol-6*).

For AD84 (N2, *pha-4–gfp*), wild-type(N2) animals were microinjected with 75 ng μl^{-1} pSP1(*pha-4–gfp*) and 75 ng μl^{-1} pRF4(*rol-6*). An extrachromosomal array was integrated as described[49] and outcrossed seven times.

Creation of *pha-4* constructs. All constructs were sequence verified.

pSP23: to construct the plasmid expressing tdTOMATO(RFP) driven by the *pha-4* endogenous promoter, the *gfp* transcript in the worm expression vector pPD95.77 was replaced with *tdTomato*[50]. The 2.7 kilobase sequence upstream of the *pha-4* coding region was amplified from genomic DNA by PCR and inserted upstream of *tdTOMATO* in pPD95.77.

pSP1: to construct the plasmid expressing PHA-4–GFP driven by the *pha-4* endogenous promoter, full-length *pha-4* cDNA (1,521 base pairs) was cloned from first-strand worm cDNA by PCR amplification and inserted in frame and upstream of the *gfp* sequence in the worm expression vector pPD95.77. The 2.7 kb sequence upstream of the *pha-4* coding region was amplified from genomic DNA by PCR and inserted upstream of the *pha-4* cDNA.

pSP15: the plasmid containing untagged PHA-4 driven by the endogenous *pha-4* promoter was constructed by amplifying the *pha-4* promoter and cDNA sequence from pSP1 by PCR amplification and inserting it in place of the *gfp* sequence in pPD95.77. A stop codon was added to the 3′ end of the *pha-4* cDNA sequence by PCR.

Lifespan analyses. Lifespan analyses were performed as described previously[35]. All lifespan analyses were conducted at 20 °C unless otherwise stated. JMP IN 5.1 software was used for statistical analysis to determine means and percentiles. In all cases, *P*-values were calculated using the log-rank (Mantel–Cox) method. Lifespan per cent decreases were determined by dividing the shorter lifespan by the longer lifespan, subtracting 1, and multiplying by 100. Mean lifespan data were used for per cent decreases.

RNA isolation, semi-quantitative RT–PCR and quantitative RT–PCR. Total RNA was isolated from synchronized populations of approximately 10,000 day-1 reproductive adults. Total RNA was extracted using TRIzol reagent (GIBCO). cDNA was created using the Quantitec Reverse Transcriptase kit (Qiagen). For semi-quantitative RT–PCR, serial dilutions of 5×, 10× and 20× were used for PCR reactions. For each primer pair, cycle times and primer concentrations were optimized to ensure linear amplification. Quantification was performed on 10× dilution reactions using Gel-Doc software and levels were normalized to *act-1* cDNA. SybrGreen real-time Q-PCR experiments were performed as described in the manual using ABI Prism79000HT (Applied Biosystems) and cDNA at a 1:20 dilution. All Q-PCR experiments were normalized to *act-1* mRNA levels.

GFP localization. GFP localization analysis was performed as described previously[13]. Worms were grown on OP50 *Escherichia coli* and images were taken on day 1 of adulthood unless otherwise noted. GFP is shown in green and is merged with differential interference contrast images. Black and white images represent only the GFP channel with fluorescence shown in white.

Pumping rate assays. Pumping rates of wild-type (N2) worms and *eat-2 (ad1116)* mutant worms on various RNAi bacteria were determined by counting pumps of the terminal pharyngeal bulb for one-minute intervals to determine pumps per min. The pumping rates of ten worms per condition were determined and averaged to determine the rates represented in the text. Worms were synchronized by transferring and growing eggs on empty vector RNAi until day 1 of adulthood. On day 1 of adulthood, worms were transferred to RNAi treatments and pumping rates were determined after forty-eight hours of RNAi treatment. Functionality of RNAi was determined by examining AD84 (*pha-4–gfp*) animals using fluorescence microscopy (Supplementary Fig. 5).

Bacterial dietary restriction (BDR) lifespan analysis. Synchronized populations of eggs were hatched and grown at 25 °C on NG agar plates containing OP50 *E. coli* until day 1 of adulthood. At day 1 of adulthood, animals were transferred to new plates of OP50 *E. coli* containing FUDR at 100 μg ml^{-1} and shifted to 20 °C. At day 2 of adulthood, worms were transferred into liquid culture (discussed below) and placed on a gentle rocker at 15 °C for the remainder of the lifespans. Liquid cultures were done in 12-well cell culture plates containing 1 ml of culture per well. Each lifespan consisted of 4 wells with 15 worms per well ($n = 60$). Lifespans were scored and worms transferred to new cultures every 3–4 days.

Liquid cultures were prepared using an overnight culture of OP50 *E. coli* grown at 37 °C. Bacteria were washed three times in S-Basal medium containing 5 μg ml^{-1} cholesterol, carbenicilln (50 μg ml^{-1}), tetracycline (1 μg ml^{-1}) and kanamycin (10 μg ml^{-1}). The bacterial concentration was adjusted to 1.5×10^9 cells ml^{-1} in S-Basal medium containing cholesterol, carbenicillin, tetracycline and kanamycin at the concentrations above. Serial dilutions were performed to achieve bacterial concentrations of 7.5×10^8, 1.5×10^8, 7.5×10^7, 2.5×10^7 and 5×10^6 cells ml^{-1}. Cultures contained FUDR at 100 μg ml^{-1} for the first twelve days of lifespan analysis to block worm reproduction.

Temperature-sensitive inactivation of *pha-4(zu225);smg-1(cc546ts).* *pha-4(zu225);smg-1(cc546ts)*[22] double mutant worms are grown at 25 °C to inactivate *smg-1* and allow functional *pha-4* to be made. *pha-4* was inactivated by shifting double mutants to 15 °C, restoring *smg-1* activity, which results in degradation of the *pha-4(zu225)* allele, after the first day of adulthood, thus avoiding any developmental defects owing to loss of *pha-4* during larval stages. All control worms were treated identically.

47. Brenner, S. The genetics of *Caenorhabditis elegans. Genetics* **77,** 71–94 (1974).
48. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10,** 3959–3970 (1991).
49. Hope, I. A. C. *elegans: A Practical Approach* (ed. Hope, I. A.) 88–90 (Oxford University Press, Oxford, 1999).
50. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature Biotechnol.* **22,** 1567–1572 (2004).

ARTICLES

Prion recognition elements govern nucleation, strain specificity and species barriers

Peter M. Tessier[1] & Susan Lindquist[2]

Prions are proteins that can switch to self-perpetuating, infectious conformations. The abilities of prions to replicate, form structurally distinct strains, and establish and overcome transmission barriers between species are poorly understood. We exploit surface-bound peptides to overcome complexities of investigating such problems in solution. For the yeast prion Sup35, we find that the switch to the prion state is controlled with exquisite specificity by small elements of primary sequence. Strikingly, these same sequence elements govern the formation of distinct self-perpetuating conformations (prion strains) and determine species-specific seeding activities. A Sup35 chimaera that traverses the transmission barrier between two yeast species possesses the critical sequence elements from both. Using this chimaera, we show that the influence of environment and mutations on the formation of species-specific strains is driven by selective recognition of either sequence element. Thus, critical aspects of prion conversion are enciphered by subtle differences between small, highly specific recognition elements.

The ability of proteins to form β-sheet-rich amyloids is associated not only with disease[1], but also with diverse normal biological functions, including cell adhesion[2], skin pigmentation[3], adaptation to environmental stresses[4–6] and perhaps even long-term neuronal memory[7]. Prions are an unusual class of amyloid-forming proteins, whose conformations are self-templating (self-seeding) and thereby infectious. The conformationally converted prion state can be transmitted from cell to cell within or, in some cases, between organisms. Prions, too, can be either deadly or beneficial[4–6,8].

The first identified prion protein was PrP, whose conversion to the prion conformer (PrP^{Sc}) is associated with several fatal neurodegenerative diseases[8]. More recently, several prions in yeast and other fungi have been identified that are unrelated to PrP or one another[9–12]; some of these may have beneficial effects[4–6]. The most well-studied is Sup35 (refs 13, 14), a translation-termination factor whose conversion to the prion state reduces its activity. This increases the read-through of stop codons, revealing hidden genetic variation and creating complex new phenotypes in a single step[4–6].

Sup35 prions exhibit two of the most baffling aspects of prion biology that were initially identified for mammalian PrP. First, both Sup35 and PrP can adopt not just one prion conformation, but a suite of related yet structurally distinct conformations (known as strains or variants)[15–21]. Each conformation self-perpetuates and confers a distinct biological phenotype. Second, the transmission of the prion state between proteins of different species is limited by a species barrier that can occasionally be traversed[22–33]. In both yeast and mammals the ability to establish and overcome species barriers is, in some unknown way, related to the ability of prions to form distinct strains[8,13,24,31,33–37].

The carboxy-terminal domain of Sup35 encodes the translation-termination function[13,14]. Whether Sup35 exists in either a prion or a non-prion state is controlled by the interplay of two other domains[13,14]. The middle region (M) has a strong solubilizing activity and is very rich in charged residues. The amino (N) terminus is extremely amyloidogenic and of unusually low sequence complexity, composed primarily of glutamine, asparagine, glycine and tyrosine residues (Supplementary Fig. 1).

In its non-prion state NM is compact, but molten, rapidly fluctuating through diverse conformations[38]. The structure of NM in its prion state is heavily debated. Several lines of evidence suggest that two discrete regions of the N domain are in self-contact within NM fibres[39]; the region between them is sequestered from intermolecular contacts, whereas elements proximal and distal to the contacts are not part of the amyloid core. Cross-linking NM molecules at one of the intermolecular contacts, but not elsewhere, accelerates nucleation[39]. However, other lines of evidence suggest that most residues of the N domain are in intermolecular contact, stacking in-register on themselves[40–43]. Genetic evidence is also in conflict. Single substitution mutations in certain regions of the N terminus can have profound effects on many aspects of prion biology: they can inhibit replication[44,45], bias prion conversion towards the production of distinct strains[36,45], and increase or decrease the ability of prions to cross species barriers[36]. These studies would indicate that precise features of amino acid sequence have critical roles in Sup35 prion biology. Remarkably, however, scrambling the sequence of N does not prevent prion formation[42,43]. This argues that NM prion formation is mainly dependent on the amino acid composition and largely independent of primary sequence[42,43].

To shed light on these paradoxes, and to bring new technology to bear on the baffling relationship between prion strains and species barriers, we have employed surface-bound arrays of prion peptides. We find that small elements of the *Saccharomyces cerevisiae* Sup35 sequence govern prion recognition with extraordinary specificity. We exploited the properties of such sequences to identify critical prion recognition elements *de novo* in a distantly related yeast species. Remarkably, the recognition elements identified using peptide arrays

[1]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. [2]Howard Hughes Medical Institute, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA.

govern nucleation, the formation of distinct prion strains and the capacity of prions to cross species-specific transmission barriers.

Identification of prion recognition elements

We sought to determine how natively unfolded non-prion conformers achieve self-recognition. Is it encoded by specific sequence elements? Or is it a distributed property of the N domain's unusual low-complexity sequence? To investigate this question we sought to interrogate the affinity of full-length proteins for diverse short elements of the NM sequence. However, short peptides from aggregation-prone proteins are often poorly soluble, precluding assessment of such interactions in solution.

To overcome this problem, we arrayed a library of overlapping denatured 20-mer peptides derived from the sequence of the prion domain of *S. cerevisiae* Sup35 (ScNM) on glass slides. An extensive library of 136 overlapping peptides from ScNM was synthesized. Each peptide carried 20 residues of prion sequence at its C terminus, preceded by a 14-carbon hydrophilic spacer and an N-terminal, double lysine tag for covalent immobilization. The peptides, which should occupy ~2–3 nm in a random coil conformation under these conditions, were printed with an average spacing of ~10–12 nm to restrict self-interactions. Each peptide sequence was shifted from that of the preceding peptide by one to six residues. In the final version of the arrays smaller intervals were employed to provide higher resolution in critical regions that were identified in preliminary experiments. (The density of coverage is indicated by the spacing of the bars in the figures, with the *x*-axis referencing the centre residue of the peptide.) The ability of the immobilized peptides to interact with soluble, full-length NM was examined by incubating fluorescently labelled protein with the peptide arrays. The label was introduced via a cysteine substitution mutation (alanine to cysteine at residue 230, A230C) located in a region that lies far from the residues involved in prion formation[39,44].

Incubating full-length ScNM with the peptide arrays for two hours led to a strong accumulation of label over only a very small set of peptides: those encompassing amino acids 9–28, 10–29, 11–30, 12–31 and 20–39 (Fig. 1a, b and Supplementary Table 2). These interactions were highly reproducible (Fig. 1a). Strikingly, these sequences lie within the region previously found by mutational analysis[44] and crosslinking[39] to strongly influence ScNM prion assembly. They also overlap with one of two regions previously identified as being in self-contact within mature fibres[39].

To determine if other peptide regions could interact with ScNM, albeit with lower efficiency, we took advantage of the fact that the spontaneous assembly of the full-length protein, even at fivefold higher concentrations, is very slow in quiescent reactions[46]. At this concentration and with a higher fraction of the protein carrying the fluorescent probe (75% versus 5% of protein), label could be detected at a second set of peptides, spanning residues ~90–120 after one to two days (Fig. 1c, Supplementary Table 2 and data not shown). This region corresponded to the second previously identified site of intermolecular contact within mature ScNM fibres (~85–105)[39]. The reactivity of the surface-bound peptides indicates that these regions are not only sites of intermolecular contact in mature fibres[39], but also represent highly specific self-recognition elements within soluble molten, non-prion conformers.

Recognition elements nucleate NM amyloids

What is the nature of the protein bound to the peptide arrays? The fluorescence detected at specific peptide spots after two hours (Fig. 1d) continued to increase after even a full day of incubation (Fig. 1e). In a further attempt to saturate peptide reactivity on the arrays, the slides were pre-incubated with unlabelled ScNM for five days. However, when subsequently incubated with labelled protein, these spots continued to accumulate label as above (Fig. 1f).

These results indicated that the growing signals at specific spots might not simply be due to slow recognition of rare soluble conformers. Rather, they might represent sites of prion assembly. Indeed, the fluorescent signal was resistant to extensive washing with 2% SDS, a characteristic of NM amyloids[46]. Further, when the material accumulating at these specific spots was scraped from the slide and imaged by transmission electron microscopy, abundant fibres with the characteristic diameter of NM prion fibres[47] were observed (Fig. 1g).

The concentration of peptides in the arrayed spots is ~300-fold greater than the protein in bulk solution. Thus, soluble proteins might achieve critical nucleating contacts with peptides on the arrays more rapidly than with other full-length proteins in solution. However, the striking correspondence between the peptides at which fibres accumulated and the amino-acid segments that form intermolecular contacts in mature ScNM amyloids[39] suggests an alternative explanation: the peptides might capture ScNM amyloids that assemble in solution during the incubation. To test this, labelled ScNM fibres were sonicated into small fragments and incubated with the arrays. The fluorescence that accumulated at peptide spots was broadly distributed and similar to background (data not shown). We conclude that the fibres associated with specific peptide spots assembled there *in situ* from soluble NM. It was not possible to determine whether the interacting soluble species is monomeric or oligomeric. Regardless, our data establish that with an extraordinary

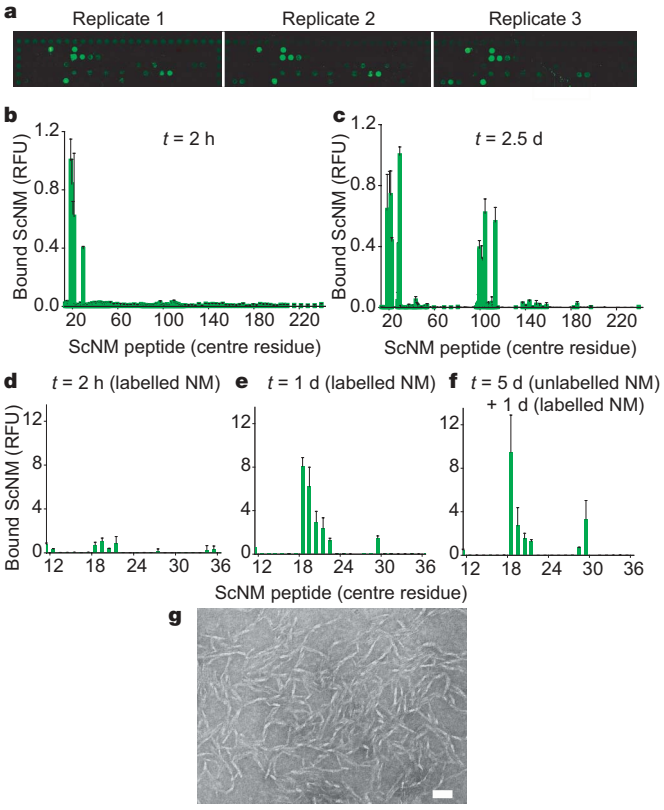


Figure 1 | Identification of recognition sequences within ScNM using peptide arrays. **a**, Image of a triplicate array of sequential overlapping 20-mer ScNM peptides after incubation with labelled full-length ScNM. **b–c**, Quantification of the fluorescence of labelled full-length ScNM bound to a similar peptide array after two hours (**b**; 1 μM, 5% Alexa Fluor 555) and two and a half days (**c**; 5 μM, 75% Alexa Fluor 555). The relative fluorescence intensity (RFU) for each 20-mer peptide is displayed at its central residue on the *x* axis. **d–f**, Quantification of the fluorescence of labelled full-length ScNM (1 μM, 5% Alexa Fluor 555) bound to ScNM peptides after two hours (**d**), one day (**e**) and one day after preincubation with unlabelled full-length ScNM (1 μM) for five days (**f**). **g**, Transmission electron micrograph of full-length ScNM fibres that assembled on a peptide array after two and a half days. The scale bar is 50 nm. For **b** and **c**, the fluorescence at each time point was normalized to the maximum fluorescence at that time point (set at 1.0). For **d–f**, fluorescence was normalized to the maximum value in **d** (set at 1.0). All fluorescence values are reported as median + s.d.

degree of specificity small elements of the Sup35 prion sequence are able to recognize molten conformers of full-length NM and convert them via this recognition to a self-templating state.

Candida albicans NM behaves in a similar manner

Next, we asked if another prion protein might have specific, highly localized recognition elements and if our method could be used to identify them. We employed the homologous prion domain of Sup35 from *C. albicans*, a species separated from *S. cerevisiae* by >800 million years[48]. The sequence of *C. albicans* NM (CaNM) shares no significant tracts of alignable sequence with *S. cerevisiae* NM[23], but is of similar low complexity and amino acid composition (Supplementary Fig. 1). CaNM was fluorescently labelled and incubated with a library of 128 overlapping CaNM peptides. After two hours, CaNM reacted in a highly-specific manner with a small cluster of peptides: amino acids 59–78, 60–79, 61–80, 62–81, 63–82, 66–85 and 67–86 (Fig. 2a and Supplementary Table 3). After longer incubations (one to two days) at fivefold higher CaNM concentrations, and with a higher fraction of the protein carrying the fluorescent probe (75% versus 5% of protein), label accumulated at a second set of peptides, spanning residues ~110–130 (inset of Fig. 2a, Supplementary Table 3 and data not shown). As with ScNM, the signals obtained with CaNM were not saturable and continued to increase over several days of incubation (data not shown). Further, the signals were resistant to extensive washing with 2% SDS. Remarkably, for two distantly related yeast species, short peptides from two distinct regions of the prion domain display a similar pattern of recognition by the full-length soluble proteins and a similar capacity to convert them, via this recognition, to a self-templating state.

Reconstitution of the species barrier on arrays

Prion fibres of *S. cerevisiae* and *C. albicans* Sup35 can seed the polymerization of their own proteins, but a species barrier prevents efficient cross-seeding[22,23,34,36]. The striking specificity by which specific peptides interact with and nucleate assembly of full-length ScNM and CaNM prompted us to ask if the species barrier could be explained by the recognition elements uncovered by the peptide arrays. Indeed, when labelled, full-length ScNM was incubated with a peptide array displaying 128 CaNM peptides for two hours, no association was detected above background (Fig. 2b). The interaction of ScNM with peptides in its own recognition element (encompassing residues 9–39) was at least 30 fold greater than its interaction with any of the CaNM peptides. Conversely, incubating labelled full-length CaNM with an array displaying 136 ScNM peptides produced only background levels of fluorescence (Fig. 2c). The interaction of CaNM with its own peptides (encompassing residues 59–86) was 28-fold greater than for any of the ScNM peptides. This high degree of specificity was maintained even after two days (data not shown).

To test the strength of the species barrier observed on the peptide arrays more stringently, we mixed proteins of both species, each labelled with a different colour dye (ScNM, green; CaNM, red), and incubated them with arrays containing both peptides. The NM proteins of each species interacted only with a small subset of peptides from their own sequences (Fig. 2d).

Next, we employed a promiscuous *S. cerevisiae*/*C. albicans* NM chimaera that has been shown previously to traverse the species barrier between *S. cerevisiae* and *C. albicans*[22,36]. This chimaeric protein contains segments from both ScNM and CaNM (residues: *S. cerevisiae*, 1–40 and 124–253; *C. albicans*, 49–141; Supplementary Fig. S1)[22,36]. Incubating the full-length NM chimaera with an array displaying libraries of both ScNM and CaNM peptides revealed that it was able to interact with the prion recognition elements from both species in a highly specific manner (Fig. 2e).

Formation of species-specific prion strains

These results suggest that (1) the species barrier between ScNM and CaNM can be defined in terms of the differential affinities of soluble protein for specific short sequence elements that have the capacity to initiate a self-templating state, and (2) crossing this barrier is based on the ability of soluble protein to interact promiscuously with sequence elements from both species. To test these hypotheses more rigorously, we exploited the ability of the Sc/Ca NM chimaera to assemble into distinct strains that have species-specific seeding activities[22,34,36]. Protein fibres assembled from the chimaeric protein at 15 °C seed amyloid formation of ScNM but not CaNM[36]. Fibres assembled at 37°C have the opposite specificity[36]. For ease of manipulation we confirmed that the species-specific strain properties obtained at 15 °C were also obtained at 4 °C at the protein concentrations employed for our arrays (data not shown).

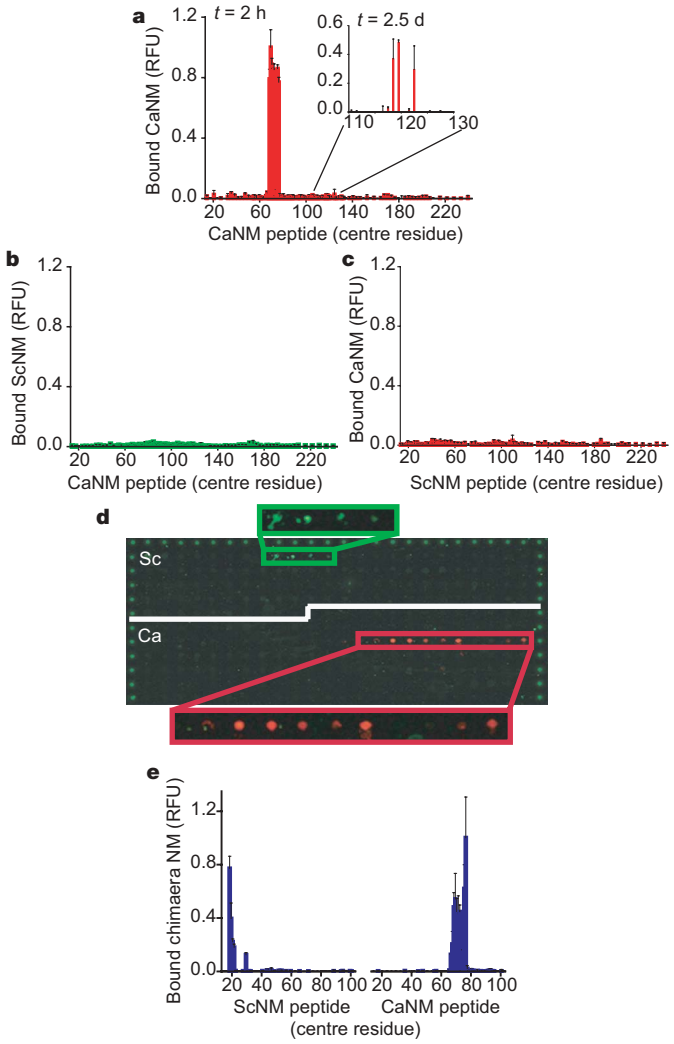


Figure 2 | Analysis of CaNM recognition sequences and the species barrier between *S. cerevisiae*/*C. albicans* NM. **a**, Quantification of the fluorescence of labelled full-length CaNM bound to overlapping 20-mer CaNM peptides after two hours (1 μM, 5% Alexa Fluor 647) and two and a half days (inset, 5 μM, 75% Alexa Fluor 647). The fluorescence at each time point was normalized to the maximum fluorescence at that time point (set at 1.0). **b**, Quantification of the fluorescence of labelled full-length ScNM (1 μM, 5% Alexa Fluor 555) bound to overlapping 20-mer CaNM peptides after two hours. **c**, Quantification of the fluorescence of labelled full-length CaNM (1 μM, 5% Alexa Fluor 647) bound to overlapping 20-mer ScNM peptides after two hours. **d**, Image of a single peptide array that was simultaneously incubated with labelled full-length ScNM (green, Alexa Fluor 555) and CaNM (red, Alexa Fluor 647) for two hours (1 μM final concentration of each protein, 5% label). **e**, Quantification of the fluorescence of labelled full-length Sc/Ca NM chimaera (1 μM NM, 5% Alexa Fluor 647) bound to both ScNM and CaNM peptides after two hours of incubation. All fluorescence values are reported as median + s.d.

As described above, when the chimaeric protein was incubated with the peptide arrays at 25 °C, it interacted with peptides from both species (Fig. 2e). However, at 37 °C it interacted selectively with CaNM peptides (Fig. 3a). At 4 °C it interacted selectively with ScNM peptides (Fig. 3b). Thus, the ability of the chimaeric protein to assemble into distinct species-specific strains at different temperatures is enciphered by the same small sequence elements that nucleate amyloid assembly.

Next, we tested the hypothesis that the effects of mutations on the formation of species-specific strains could be explained by these same prion recognition elements. As previously reported, changing serine residue 17 to arginine (S17R) in the Sc/Ca chimaera favours assembly of a prion strain that selectively seeds CaNM[36]. Conversely, changing four glycines at positions 70, 71, 80 and 81 to alanines (4G/A) favours assembly of a strain that selectively seeds ScNM[36]. When the S17R chimaera was incubated with the arrays at 25 °C, binding to all ScNM peptides was reduced to background (Fig. 3c). Importantly, binding to CaNM peptides was unaffected (Fig. 3c). Similarly, when the 4G/A chimaera was incubated with the arrays, binding to all but one of the CaNM peptides was greatly reduced but binding to ScNM peptides was similar to the original chimaera (Fig. 3d).

Mechanism of prion recognition

How do the mutations in the Sc/Ca chimaera alter prion recognition and strain formation? They might bias the conformations sampled by the molten full-length proteins such that particular recognition elements are masked. Alternatively, they might directly interfere with interactions between the full-length proteins and their cognate recognition elements. To investigate this question, the ability of the original chimaeric protein to interact with arrays containing mutant peptides was tested. Labelled chimaeric protein bound robustly to the wild-type ScNM and CaNM peptides (dark blue bars, Fig. 4a and b). However, it exhibited no interaction above background with any ScNM peptides containing the S17R mutation (yellow bars, Fig. 4a) or any CaNM peptides containing the 4G/A mutations (pink bars, Fig. 4b). Because the original chimaera reacted with wild-type peptides on the same array, it must have displayed its own recognition element. Its inability to interact with the mutant peptides, therefore, indicates that the mutations directly disrupt the recognition function of the sequence elements rather than solely altering the conformations of the soluble protein. Thus, these mutations, which bias prions toward the formation of distinct strains and alter cross-species prion transmission, do so by directly interfering with recognition of the prion specificity elements.

Discussion

Using arrays of surface-bound peptides, we find that small elements of prion sequence govern self-recognition with exquisite specificity. These recognition elements are alone sufficient to drive the conversion of full-length NM in the non-prion conformation—molten,

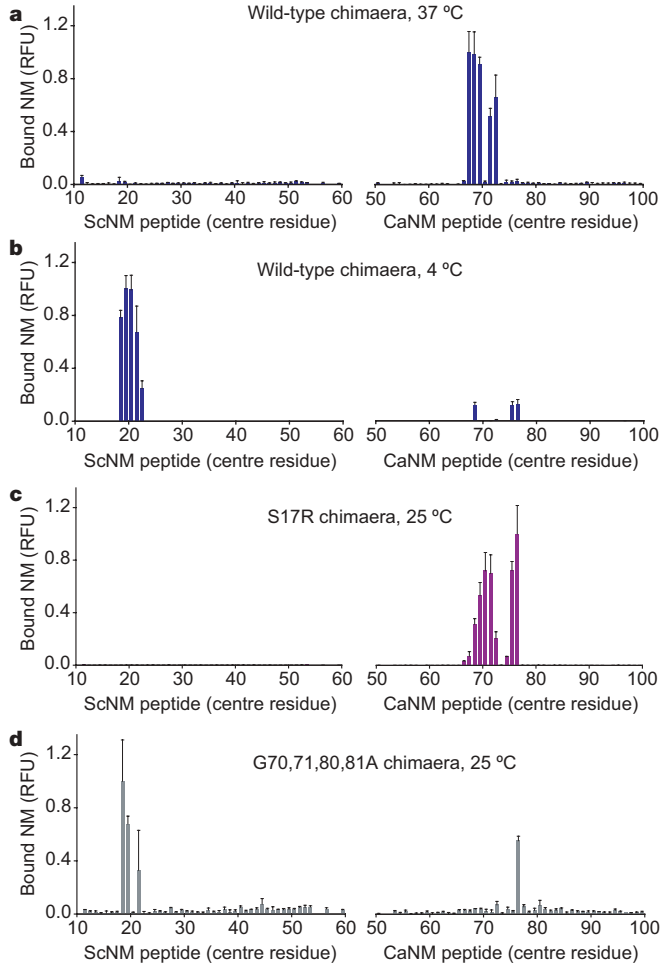


Figure 3 | Analysis of the conformational preference of the *S. cerevisiae*/*C. albicans* NM chimaera. **a**–**d**, Quantification of the relative binding of various labelled full-length NM chimaeric proteins to overlapping 20-mer ScNM and CaNM peptides: **a**, Sc/Ca chimaera at 37 °C; **b**, Sc/Ca chimaera at 4 °C; **c**, S17R Sc/Ca chimaera at 25 °C; and **d**, G70, 71, 80, 81A Sc/Ca chimaera at 25 °C. The peptide arrays were incubated with each prion domain (1 μM NM, 5% Alexa Fluor 647) for two hours. All fluorescence values are reported as median + s.d.

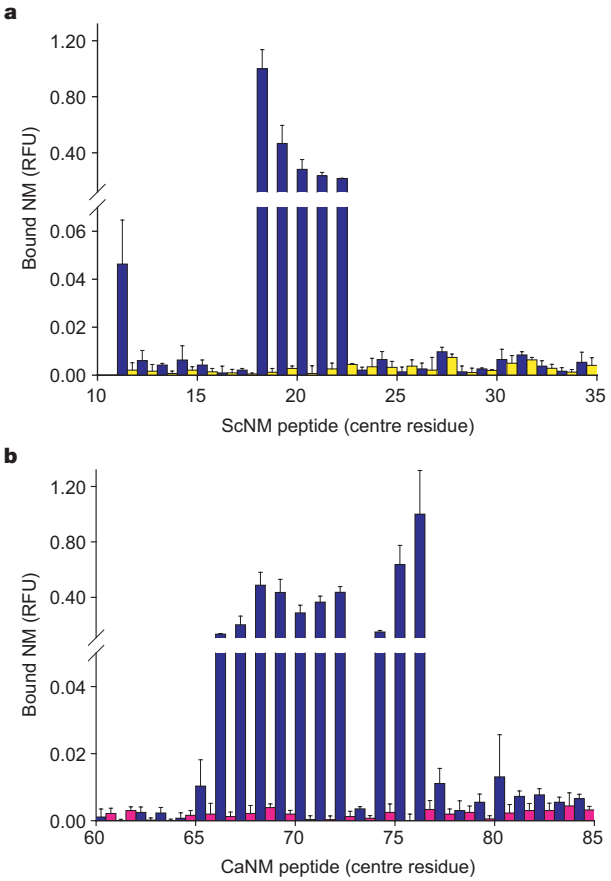


Figure 4 | Analysis of the mutational disruption of the ScNM and CaNM recognition elements. **a**, Quantification of the relative affinity of the full-length Sc/Ca NM chimaera for wild-type ScNM peptides (dark blue bars) and ScNM peptides containing the S17R mutation (yellow bars). **b**, Quantification of the relative affinity of the full-length Sc/Ca NM chimaera for wild-type CaNM peptides (dark blue bars) and CaNM peptides containing the G70, 71, 80, 81A mutations (pink bars). The peptide arrays were incubated with the NM chimaera (1 μM NM, 5% Alexa Fluor 647) for two hours. All fluorescence values are reported as median + s.d.

rapidly fluctuating, natively unfolded conformers—into self-templating amyloid conformations. Moreover, it is these same sequence elements that govern the formation of distinct prion strains and determine transmission barriers between species.

The highly localized and species-specific nature of the interactions detected with peptide arrays was unexpected for several reasons. For example, when the sequence of the *S. cerevisiae* Sup35 prion domain is scrambled, the protein still forms prions[42,43]. This had suggested that prion formation was a distributed property, owing to the unusual amino acid composition of the N domain, rather than a localized property of specific sequences within it. Indeed, scrambling of the prion domain sequence has occurred in nature for different yeast species. Although the *S. cerevisiae* and *C. albicans* N domains are both of low complexity (half of their residues are glutamines or asparagines and most others are tyrosines or glycines; Supplementary Fig. 1), their amino acid sequences cannot be aligned[23]. Yet CaNM forms prions with a similar efficiency as ScNM[22,23]. This, too, might suggest that prion formation is a distributed property governed primarily by amino acid composition.

However, full-length ScNM and CaNM interact with only a small number of peptides derived from their own sequences and do so in a highly localized and specific manner. But how can our data be reconciled with the ability of scrambled sequences to produce bona fide heritable prion elements? One explanation is that the scrambled NM sequences may form prions at very low efficiency; they were expressed at very high levels and their rates of prion conversion relative to the wild-type were not reported[42]. Alternatively, it may be that scrambling N domain sequences frequently results in the formation of new recognition elements. Given that the prion state produces many changes in phenotype (some detrimental and some beneficial)[4–6], the efficiency of prion conversion may have been conserved for ScNM and CaNM through selective pressures, despite sequence variation.

The peptides in the recognition elements we identified also have the unique capacity to drive molten non-prion conformers into a self-templating state. It has recently been reported that small amyloidogenic peptides from the N domain of ScNM (residues 7–12 and 7–13, which partially overlap with one of our recognition elements) can form microcrystals that yield high-resolution structures[49]. Our identification of critical nucleating peptides, therefore, holds promise for future high-resolution analysis of the bona fide templating structures of the yeast prion [PSI^+], and potentially the comparative analysis of such structures from organisms separated by 800 million years of evolution.

Remarkably, the non-prion conformers of both ScNM and CaNM interacted initially and most strongly with a small cluster of their own peptides towards the N-terminal side of the N domain (Figs 1b and 2a). Additionally, both proteins could interact with a second discrete cluster of their own peptides towards the C-terminal end of the N domain, but only after much longer incubations and at higher protein concentrations (Figs 1c and 2a, inset). This finding indicates that the nature of prion nucleation, conformational conversion and templating for Sup35 may have been conserved during evolution. If true, this would provide strong evidence for our previous hypothesis that the phenotypic variation created when Sup35 switches to the prion state is of considerable biological importance[4,5].

It is striking that the specific recognition sites identified on the arrays correspond to regions of intermolecular association identified in ScNM fibres by other methods[39]. Although our results cannot distinguish between proposed structural models for ScNM prion fibres[39–41], the data seem most consistent with one of them[39]: that two distinct regions in the N domain are in self-contact, with the intervening region sequestered from intermolecular contacts, and regions distal and proximal to the contacts being outside the amyloid core[39]. They are also consistent with cross-linking data suggesting that one of the two contact sites has a dominant role in nucleation[39].

Our data also suggest a new framework for conceptualizing the formation of species barriers and prion strains. Species barriers are known to depend, in part, on the primary sequence of prions[22–33]. We find that they are determined by highly localized elements of amino acid sequence. Moreover, these are the very same elements that control nucleation. Species barriers are also known to be influenced by prion strains[8,13,24,31,33–37], but the molecular connection has been elusive. For the promiscuous Sc/Ca NM chimaera the remarkable concordance between the way that specific mutations and environmental conditions determine the particular peptides that can nucleate soluble chimaeric protein on the arrays (Fig. 3) and the way they determine the species-specific propagation of different chimaeric strains[36] suggests the following relationship—within the same region of the same protein, small elements of primary sequence can initiate distinct intermolecular recognition events that drive the formation of different amyloid structures. These same sequence elements are the intermolecular contact sites that are then involved in templating those structures to soluble protein for propagation of the prion strain.

From a methodological perspective, it is empowering that peptide arrays, which are typically used to study the specificities of folded proteins, can also be used to identify sequences that drive structural transitions in natively unfolded proteins. This method has the potential to illuminate many difficult problems in human disease biology related to the conformational conversion of amyloidogenic proteins. Peptide arrays might also be employed in high-throughput drug screens for compounds that prevent protein assembly, promote it, or redirect it to alternative forms. Finally, the fact that highly ordered amyloid fibres can be specifically nucleated and assembled at patterned sites on solid surfaces, and that this assembly process can be tightly controlled by simple changes in environmental conditions (for example, temperature), may have important implications for fabricating materials and devices with complex, nanostructured features.

Perhaps the most important implication of our work is that the very nature of protein folding for Sup35 prions is profoundly different from that of proteins with globular domains. The latter is typically governed by a large number of intramolecular interactions that collectively and cooperatively drive folding of the entire domain. In contrast, we find that the folding of two distantly related yeast prion proteins into amyloid fibres of defined tertiary structure is controlled by intermolecular interactions between very small elements of primary sequence. Peptide arrays offer an opportunity to determine whether other amyloidogenic proteins fold by a similar mechanism.

METHODS SUMMARY

ScNM and CaNM contained a C-terminal 7×His-tag, as did the Sc/Ca NM chimaera constructs except for S17R, which contained a 6×His-tag. The NM cysteine mutants at or near the C terminus (*S. cerevisiae* A230C, *C. albicans* S227C and all chimaeras at the extreme C terminus) were labelled with maleimide-functionalized Alexa Fluor 555 or 647 (Invitrogen).

Each peptide contained a double alanine tag at its N terminus, a 20-residue segment derived from the Sup35 sequences, a hydrophilic linker (1-amino-4,7,10-trioxa-13-tridecanamine succinimic acid[50]) and a double lysine tag at its C terminus. The peptides were then printed onto hydrogel glass slides (Nexterion Slide H, Schott) functionalized with reactive *N*-hydroxysuccinimide (NHS) ester moieties. The NM proteins were denatured in 6 M guanidine hydrochloride at 100 °C, and then diluted 125-fold to a final concentration of 1–5 μM NM and a label-per-protein molar ratio of 5–75%. Each peptide array was incubated individually with 2–3 ml of diluted NM without mixing. The peptide arrays were then washed extensively with 2% SDS, water and methanol, and then spun dry. The arrays were then imaged using a Genepix 4000A scanner and the median fluorescence values of two or three replicates for the peptide spots were quantified using Genepix Pro 6.0 software (Molecular Devices).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 March; accepted 16 April 2007.
Published online 9 May 2007.

1. Selkoe, D. J. Folding proteins in fatal ways. *Nature* **426**, 900–904 (2003).

2. Chapman, M. R. *et al.* Role of *Escherichia coli* curli operons in directing amyloid fiber formation. *Science* **295,** 851–855 (2002).
3. Fowler, D. M. *et al.* Functional amyloid formation within mammalian tissue. *PLoS Biol.* **4,** 100–107 (2006).
4. True, H. L. & Lindquist, S. L. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407,** 477–483 (2000).
5. True, H. L., Berlin, I. & Lindquist, S. L. Epigenetic regulation of translation reveals hidden genetic variation to produce complex traits. *Nature* **431,** 184–187 (2004).
6. Eaglestone, S. S., Cox, B. S. & Tuite, M. F. Translation termination efficiency can be regulated in *Saccharomyces cerevisiae* by environmental stress through a prion-mediated mechanism. *EMBO J.* **18,** 1974–1981 (1999).
7. Si, K., Lindquist, S. & Kandel, E. R. A neuronal isoform of the Aplysia CPEB has prion-like properties. *Cell* **115,** 879–891 (2003).
8. Prusiner, S. B. Prions. *Proc. Natl Acad. Sci. USA* **95,** 13363–13383 (1998).
9. Wickner, R. B. & Masison, D. C. Evidence for two prions in yeast: [URE3] and [PSI]. *Curr. Top. Microbiol. Immunol.* **207,** 147–160 (1996).
10. Wickner, R. B. [Ure3] as an altered Ure2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*. *Science* **264,** 566–569 (1994).
11. Sondheimer, N. & Lindquist, S. Rnq1: an epigenetic modifier of protein function in yeast. *Mol. Cell* **5,** 163–172 (2000).
12. Coustou, V., Deleu, C., Saupe, S. & Begueret, J. The protein product of the *het-s* heterokaryon incompatibility gene of the fungus *Podospora anserina* behaves as a prion analog. *Proc. Natl Acad. Sci. USA* **94,** 9773–9778 (1997).
13. Chien, P., Weissman, J. S. & DePace, A. H. Emerging principles of conformation-based prion inheritance. *Annu. Rev. Biochem.* **73,** 617–656 (2004).
14. Tuite, M. F. & Cox, B. S. Propagation of yeast prions. *Nature Rev. Mol. Cell Biol.* **4,** 878–890 (2003).
15. King, C. Y. & Diaz-Avalos, R. Protein-only transmission of three yeast prion strains. *Nature* **428,** 319–323 (2004).
16. Tanaka, M., Chien, P., Naber, N., Cooke, R. & Weissman, J. S. Conformational variations in an infectious protein determine prion strain differences. *Nature* **428,** 323–328 (2004).
17. Bruce, M. E., McConnell, I., Fraser, H. & Dickinson, A. G. The disease characteristics of different strains of scrapie in *Sinc* congenic mouse lines: implications for the nature of the agent and host control of pathogenesis. *J. Gen. Virol.* **72,** 595–603 (1991).
18. Caughey, B., Raymond, G. J. & Bessen, R. A. Strain-dependent differences in β-sheet conformations of abnormal prion protein. *J. Biol. Chem.* **273,** 32230–32235 (1998).
19. Kocisko, D. A. *et al.* Cell-free formation of protease-resistant prion protein. *Nature* **370,** 471–474 (1994).
20. Safar, J. *et al.* Eight prion strains have PrP(Sc) molecules with different conformations. *Nature Med.* **4,** 1157–1165 (1998).
21. Derkatch, I. L., Chernoff, Y. O., Kushnirov, V. V., Inge-Vechtomov, S. G. & Liebman, S. W. Genesis and variability of [PSI] prion factors in *Saccharomyces cerevisiae*. *Genetics* **144,** 1375–1386 (1996).
22. Chien, P. & Weissman, J. S. Conformational diversity in a yeast prion dictates its seeding specificity. *Nature* **410,** 223–227 (2001).
23. Santoso, A., Chien, P., Osherovich, L. Z. & Weissman, J. S. Molecular basis of a yeast prion species barrier. *Cell* **100,** 277–288 (2000).
24. Collinge, J. Prion diseases of humans and animals: their causes and molecular basis. *Annu. Rev. Neurosci.* **24,** 519–550 (2001).
25. Chernoff, Y. O. *et al.* Evolutionary conservation of prion-forming abilities of the yeast Sup35 protein. *Mol. Microbiol.* **35,** 865–876 (2000).
26. Kushnirov, V. V., Kochneva-Pervukhova, N. V., Chechenova, M. B., Frolova, N. S. & Ter-Avanesyan, M. D. Prion properties of the Sup35 protein of yeast *Pichia methanolica*. *EMBO J.* **19,** 324–331 (2000).
27. Nakayashiki, T., Ebihara, K., Bannai, H. & Nakamura, Y. Yeast [PSI+] "prions" that are crosstransmissible and susceptible beyond a species barrier through a quasi-prion state. *Mol. Cell* **7,** 1121–1130 (2001).
28. Resende, C. *et al.* The *Candida albicans* Sup35p protein (CaSup35p): function, prion-like behaviour and an associated polyglutamine length polymorphism. *Microbiology* **148,** 1049–1060 (2002).
29. Scott, M. *et al.* Transgenic mice expressing hamster prion protein produce species-specific scrapie infectivity and amyloid plaques. *Cell* **59,** 847–857 (1989).
30. Prusiner, S. B. *et al.* Transgenic studies implicate interactions between homologous PrP isoforms in scrapie prion replication. *Cell* **63,** 673–686 (1990).
31. Collinge, J. *et al.* Unaltered susceptibility to BSE in transgenic mice expressing human prion protein. *Nature* **378,** 779–783 (1995).
32. Supattapone, S. *et al.* Prion protein of 106 residues creates an artificial transmission barrier for prion replication in transgenic mice. *Cell* **96,** 869–878 (1999).
33. Bruce, M. *et al.* Transmission of bovine spongiform encephalopathy and scrapie to mice: strain variation and the species barrier. *Phil. Trans. R. Soc. Lond. B* **343,** 405–411 (1994).
34. Tanaka, M., Chien, P., Yonekura, K. & Weissman, J. S. Mechanism of cross-species prion transmission: An infectious conformation compatible with two highly divergent yeast prion proteins. *Cell* **121,** 49–62 (2005).
35. Hill, A. F. *et al.* The same prion strain causes vCJD and BSE. *Nature* **389,** 448–450 (1997).
36. Chien, P., DePace, A. H., Collins, S. R. & Weissman, J. S. Generation of prion transmission barriers by mutational control of amyloid conformations. *Nature* **424,** 948–951 (2003).
37. Collinge, J., Sidle, K. C., Meads, J., Ironside, J. & Hill, A. F. Molecular analysis of prion strain variation and the aetiology of 'new variant' CJD. *Nature* **383,** 685–690 (1996).
38. Mukhopadhyay, S., Krishnan, R., Lemke, E. A., Lindquist, S. & Deniz, A. A. A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. *Proc. Natl Acad. Sci. USA* **104,** 2649–2654 (2007).
39. Krishnan, R. & Lindquist, S. L. Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature* **435,** 765–772 (2005).
40. Kajava, A. V., Baxa, U., Wickner, R. B. & Steven, A. C. A model for Ure2p prion filaments and other amyloids: the parallel superpleated β-structure. *Proc. Natl Acad. Sci. USA* **101,** 7885–7890 (2004).
41. Shewmaker, F., Wickner, R. B. & Tycko, R. Amyloid of the prion domain of Sup35p has an in-register parallel β-sheet structure. *Proc. Natl Acad. Sci. USA* **103,** 19754–19759 (2006).
42. Ross, E. D., Edskes, H. K., Terry, M. J. & Wickner, R. B. Primary sequence independence for prion formation. *Proc. Natl Acad. Sci. USA* **102,** 12825–12830 (2005).
43. Ross, E. D., Baxa, U. & Wickner, R. B. Scrambled prion domains form prions and amyloid. *Mol. Cell. Biol.* **24,** 7206–7213 (2004).
44. DePace, A. H., Santoso, A., Hillner, P. & Weissman, J. S. A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion. *Cell* **93,** 1241–1252 (1998).
45. King, C. Y. Supporting the structural basis of prion strains: induction and identification of [PSI] variants. *J. Mol. Biol.* **307,** 1247–1260 (2001).
46. Serio, T. R. *et al.* Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science* **289,** 1317–1321 (2000).
47. Glover, J. R. *et al.* Self-seeded fibers formed by Sup35, the protein determinant of [PSI+], a heritable prion-like factor of *S. cerevisiae*. *Cell* **89,** 811–819 (1997).
48. Heckman, D. S. *et al.* Molecular evidence for the early colonization of land by fungi and plants. *Science* **293,** 1129–1133 (2001).
49. Nelson, R. *et al.* Structure of the cross-β spine of amyloid-like fibrils. *Nature* **435,** 773–778 (2005).
50. Zhao, Z. G., Im, J. S., Lam, K. S. & Lake, D. F. Site-specific modification of a single-chain antibody using a novel glyoxylyl-based labeling reagent. *Bioconjugat. Chem.* **10,** 424–430 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Weissman for providing the CaNM and Sc/Ca NM chimaera plasmids, M. Schutkowski for assistance in designing and preparing the peptide arrays, N. Watson for performing the transmission electron microscope imaging and members of the Lindquist laboratory for helpful discussions. This research was supported by an American Cancer Society Postdoctoral Fellowship (P.M.T.), and grants from the DuPont-MIT Alliance and the NIH. S.L. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions Experimental work was performed by P. Tessier, and the data analysis and writing were conducted by P. Tessier and S. Lindquist.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.L. (lindquist_admin@wi.mit.edu).

METHODS

Mutagenesis, protein purification and cysteine labelling. Single cysteine mutations were introduced into NM using QuikChange mutagenesis (Stratagene). All NM proteins were purified as described previously[47] except that the proteins were eluted from a Ni-NTA column using low pH instead of imidazole. The NM cysteine mutants at or near the C terminus (*S. cerevisiae* A230C, *C. albicans* S227C and all chimaeras at the extreme C terminus) were labelled overnight at room temperature with maleimide-functionalized Alexa Fluor 555 or 647 (Invitrogen) using a 5:1 to 10:1 molar ratio of label:NM, and the free label was removed using a Ni-NTA column.

Peptide array synthesis, hybridization and quantification. Peptides were synthesized on modified cellulose membranes using SPOT technology[51] (JPT Peptide Technologies GmbH). The peptides were cleaved off the membranes, freeze-dried and resuspended in buffer (40% DMSO, 5% glycerol, 55% PBS, pH 9) for printing. The peptides were then printed in triplicate onto hydrogel glass slides (Nexterion Slide H, Schott) and immobilized via reactive *N*-hydroxysuccinimide (NHS) ester moieties. (It is recommended to print each peptide 4–6 times per array for improved statistics.) Each peptide spot (250 μm in diameter) was printed with 1.5 nl of peptide solution at a concentration of approximately 2.5 μM using non-contact printing (JPT Peptide Technologies GmbH). Unreacted peptides were removed from the hydrogel slides. Slides were dried and blocked with 3% BSA in PBST for 1 h. The NM proteins were denatured in 6 M guanidine hydrochloride at 100 °C for approximately 20 min, and then diluted 125-fold in PBST containing 3% BSA to a final concentration of 1–5 μM NM and a label-per-protein molar ratio of 5–75%. Each peptide array was incubated individually with 2–3 ml of diluted NM in an Altas hybridization chamber (BD Biosciences) for the indicated times without mixing. The peptide arrays were then washed 5 times with 50 ml of 2% SDS for 30 min, 5 times with 50 ml of water, 3 times with 50 ml of methanol and then spun dry. The methanol washes were not essential but helped prevent uneven drying of the slides. The arrays were then imaged using a Genepix 4000A scanner and the median fluorescence values of two or three replicates for the peptide spots were quantified using Genepix Pro 6.0 software (Molecular Devices). The Genepix composite pixel intensity (CPI) threshold was varied between different experiments to illustrate differences in protein interaction specificity.

51. Frank, R. The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports–principles and applications. *J. Immunol. Methods* **267**, 13–26 (2002).

LETTERS

Water vapour and hydrogen in the terrestrial-planet-forming region of a protoplanetary disk

J. A. Eisner[1]

Planetary systems (ours included) formed in disks of dust and gas around young stars. Disks are an integral part of the star and planet formation process[1,2], and knowledge of the distribution and temperature of inner-disk material is crucial for understanding terrestrial planet formation[3], giant planet migration[4], and accretion onto the central star[5]. Although the inner regions of protoplanetary disks in nearby star-forming regions subtend only a few nano-radians, near-infrared interferometry has recently enabled the spatial resolution of these terrestrial zones. Most observations have probed only dust[6], which typically dominates the near-infrared emission. Here I report spectrally dispersed near-infrared interferometric observations that probe the gas (which dominates the mass and dynamics of the inner disk), in addition to the dust, within one astronomical unit (1 AU, the Sun–Earth distance) of the young star MWC 480. I resolve gas, including water vapour and atomic hydrogen, interior to the edge of the dust disk; this contrasts with results of previous spectrally dispersed interferometry observations[7,8]. Interactions of this accreting gas with migrating planets may lead to short-period exoplanets like those detected around main-sequence stars[4]. The observed water vapour is probably produced by the sublimation of migrating icy bodies[9], and provides a potential reservoir of water for terrestrial planets[10].

I obtained observations with a new grism[11] (a dispersive element consisting of a diffraction grating on the vertex of a prism) at the Keck Interferometer[12,13] that show (for the first time to my knowledge) evidence of spatially resolved gaseous spectral-line emission interior to the edge of the dusty disk around a young star, MWC 480. My observations provide a spectral dispersion of $R = 230$ at wavelengths from 2.0 to 2.4 μm, and an angular resolution of approximately 1 milli-arcsecond (1 mas ≈ 5 nano-radians), which is ~0.1 AU at the 140 parsec distance to the target (the source is located in the Taurus-Auriga star-forming complex). MWC 480 has a stellar mass of $2.3 M_{\odot}$ (where $M_{\odot}$ is the solar mass), an age of ~6 Myr, and a circumstellar disk of dust and gas in keplerian rotation[14]. Previous observations found that the dusty component of the disk extended to within ~0.5 AU of the central star[15,16], and suggested the presence of hot gaseous emission at even smaller stellocentric radii[17]. My observations confirm these suggestions, and resolve hydrogen gas and water vapour in the inner disk.

My interferometry data provide a measurement of the angular size of the circumstellar emission from MWC 480 as a function of wavelength (Fig. 1). The spectral dependence of the angular size indicates a radial temperature gradient. Because cooler material emits relatively more radiation at longer wavelengths, an outwardly decreasing temperature gradient yields an increasing fraction of emission from larger radii, and hence a larger angular size, at longer wavelengths. The temperature gradient probably arises from multiple components at different stellocentric radii[17]: warm dust at radii where temperatures approach the dust sublimation temperature, and hotter gas at smaller radii.

Figure 1 also shows more compact emission at 2.165 μm relative to the surrounding continuum. This dip corresponds to the wavelength of the Brackett gamma (Brγ) electronic transition ($n = 7 \rightarrow 4$) of hydrogen. The Brγ emission observed from MWC 480 traces very hot gas (~10^4 K) that has recently recombined from ionized hydrogen[18], presumably in the inner reaches of an accretion column.

To match the features of Fig. 1, I constructed a simple physical model and fitted it directly to the measured squared visibilities (V^2)

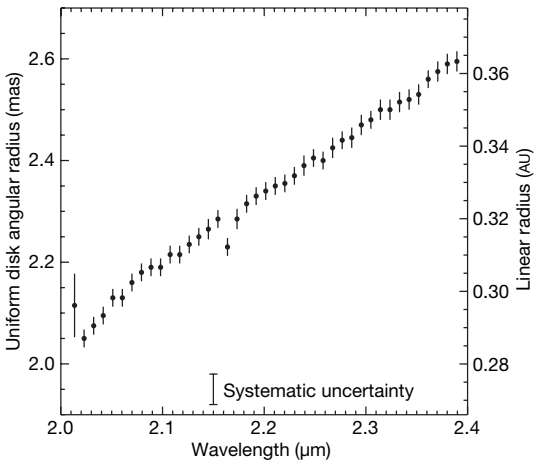


Figure 1 | Angular size of the near-infrared emission from MWC 480 as a function of wavelength. Sizes are calculated from squared visibility measurements (V^2) made with the Keck Interferometer grism on 12 November 2006. The V^2 (and fluxes; see Fig. 2), were calibrated using observations of unresolved stars of known spectral type and parallax[11]. I applied these calibrations to unresolved 'check stars' to test the calibration, and thereby estimated channel-to-channel uncertainties (1σ; standard deviation) in my flux and V^2 measurements of 3%. The shortest-wavelength data point has a larger uncertainty, because atmospheric water vapour absorption leads to lower photon counts in this channel; I estimate the uncertainty in this channel at ~10%. There is also a systematic uncertainty of approximately 5% in the overall normalization of the data, owing to errors in the absolute calibration[11]. I assume that the emission morphology is a uniform disk, for which the measured V^2 are related to the angular diameter of the source, θ_{UD}, by $V^2 = [J_1(\pi\theta_{UD}B_{proj}/\lambda)/(\pi\theta_{UD}B_{proj}/\lambda)]^2$, where B_{proj} is the projected baseline separation of the two telescopes as seen by the source, λ is the wavelength, and J_1 is a first-order Bessel function. I remove the stellar contribution to the visibilities[15], and the plotted sizes thus represent only the circumstellar emission. The angular size increases from 2.0 to 2.4 μm, indicating multiple emission components: hot gas at small stellocentric radii and warm dust at larger radii[17]. In addition, the angular size is smaller at 2.165 μm, the wavelength where hot hydrogen gas emits in the Brγ line, than at adjacent continuum wavelengths.

[1]Department of Astronomy, 601 Campbell Hall, University of California, Berkeley, California 94720, USA.

and fluxes. The model accounts for emission of the unresolved central star, circumstellar continuum emission from hot gas, spectral-line emission from hydrogen (Brγ), and continuum emission from warm dust. I take each emission component to be a uniform ring whose width is 20% of its radius. This multiple-ring model approximates a continuous disk in which dust, gas and individual gaseous atoms or molecules contribute emission at different size scales. It has previously been shown that ring models provide reasonable approximations to more complex models of inner-disk regions[15,19]. I also assume that the gaseous continuum and Brγ emission originate from stellocentric radii smaller than 0.1 AU; given the ~1 mas angular resolution of my observations, I cannot accurately distinguish between smaller radii, although I can rule out larger radii.

As indicated in Fig. 2, the model fits the data well over most of the spectral bandpass. However, there are substantial deviations between the model predictions and the data in the region between 2.0 and 2.1 μm (particularly for the fluxes). These deviations suggest the presence of hot water vapour, which has substantial opacity from 2.0 to 2.1 μm, in the inner disk of MWC 480. I include water vapour in the model by adding a ring of material with the opacity of water vapour[20]. The superior quality of the fit when water is included clearly demonstrates the presence of H_2O in the inner disk of MWC 480 (for further reasons why the model must include water vapour emission to match the data, see the Supplementary Information).

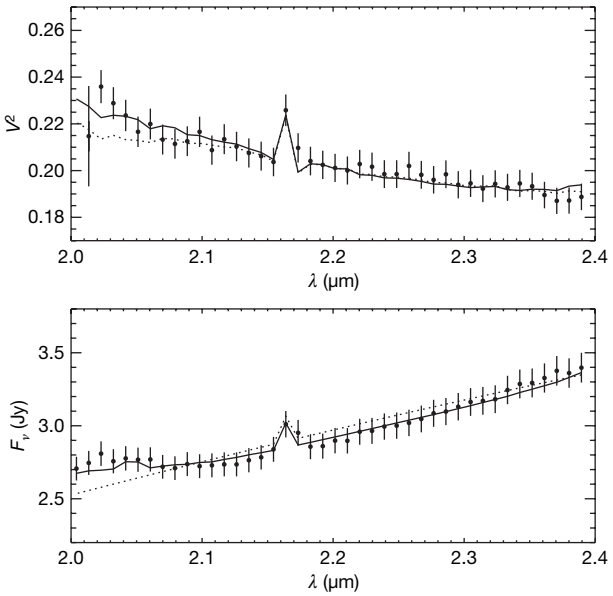


Figure 2 | Measured V^2 and fluxes, compared to the predictions of simple physical models. Plotted error bars indicate estimated 1σ (standard deviation) uncertainties of 3%; as noted above, the uncertainty in the shortest wavelength V^2 measurement is larger, approximately 10%. The dotted line shows the values computed for a model including continuum emission from warm dust and hot gas, and Brγ emission from hydrogen. This model reproduces the overall slopes seen in the data as well as the feature at 2.165 μm, but deviates from the data between 2.0 and 2.1 μm. Hot water vapour contributes significant emission at these wavelengths[20], and the solid line shows the predictions of a model that includes H_2O. Because the H_2O emission lies at a radius only slightly larger than that of the hot gaseous continuum, its effect on the visibilities is more subtle than its effect on the measured fluxes. Absorption due to terrestrial atmospheric water vapour affects the data in the shortest-wavelength channel, but not substantially in longer-wavelength channels because, in contrast to the hot water vapour in our model, water vapour at atmospheric temperature and pressure conditions lacks substantial opacity at these wavelengths[20]. The model including H_2O fits the data well over the entire observed wavelength range, and yields substantially smaller residuals between model and data than the model without water vapour. We note that the inferred sizes of the emission components in our model are smaller than the uniform disk sizes plotted in Fig. 1; this is due primarily to the different characteristic sizes for uniform disk and relatively narrower ring models[15]. Additional details of the modelling and fitting procedure can be found in previous work[17].

The best-fit model includes compact gaseous continuum emission with a temperature of (2,410 ± 125) K and Brγ emission that adds 17% ± 7% to this gaseous continuum flux at 2.165 μm; these components are situated at stellocentric radii <0.1 AU. Water vapour emission is located at a radius of (0.16 ± 0.05) AU. The temperature of this material is (2,300 ± 120) K and the column density is $(1.2 \pm 0.6) \times 10^{19}$ cm^{-2}. Dust emission originates from a radius of (0.28 ± 0.01) AU and has a temperature of (1,200 ± 60) K. Quoted 1σ uncertainties are quadrature sums of statistical uncertainties in the fits and assumed 5% systematic errors.

I illustrate the basic geometry of the innermost regions of MWC 480 in Fig. 3. As inferred from the data, gaseous emission is found at substantially smaller stellocentric radii than the dust. The inner edge of the dust disk lies at ~0.3 AU, where the disk temperature approaches the sublimation temperature of silicate dust[21]. Because I also include contributions from gas in my analysis, the inferred dust size is larger, and the temperature lower, than in previous work where only dust emission was considered[15,16]. Emission from H_2O is found at ~0.15 AU, and hot gaseous continuum and Brγ emission at still smaller radii (<0.1 AU). The observed Brγ emission probably originates in an accretion column, near the shock formed when infalling material impacts high-latitude regions of the star[5,22,23].

My results substantiate models of inner-disk structure proposed previously for similar star + disk systems[24]. Comparison with these models suggests that to reproduce the ratio of flux in the gas and dust components determined here (~0.5), a mass accretion rate of $\geq 10^{-7}\,M_\odot\,yr^{-1}$ must be invoked for MWC 480. At such high accretion rates, the inner-disk gas becomes optically thick to its own radiation, and can therefore produce relatively more flux. Furthermore, models with this high accretion rate produce strong Brγ emission, consistent with my detection. My measurement of gas interior to the edge of the dust disk also confirms previous suggestions of such structure based on modelling of spatially unresolved (but spectrally resolved) observations of gaseous emission lines[22,25,26] or very low dispersion interferometric measurements[17].

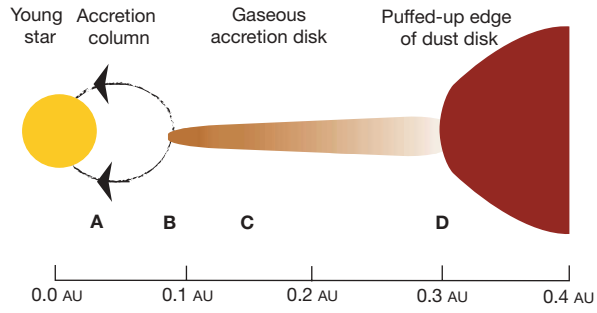


Figure 3 | The environment within 1 AU of the young star MWC 480. Dust in the inner disk around MWC 480 extends inward to approximately 0.3 AU (labelled D), where the disk temperature becomes high enough to sublimate the dust. The ring-like appearance of the dust sublimation front is compatible with my data, and with physical models predicting a puffed-up and curved inner edge[19,29]. Interior to the dust sublimation radius, gaseous material continues to accrete onto the young star (and may also be blown away in a tenuous wind[30], which is not depicted). I have portrayed a continuous gaseous disk whose optical depth increases at smaller radii (based on a sketch for a similar young star in ref. 24). I observe hot continuum emission from the hottest, densest gaseous material at stellocentric radii smaller than 0.1 AU (labelled B) and spectral-line emission from hot water vapour at slightly larger radii (labelled C). Finally, I observe Brγ emission at radii smaller than 0.1 AU, which traces hydrogen gas at high temperatures (~10^4 K). I speculate that this emission arises in the hot accretion shock where infalling gas strikes the stellar surface (labelled A). The ~1 mas angular resolution of my observations translates into a spatial resolution of approximately 0.1 AU at the distance to MWC 480.

The finding that gaseous emission, and Brγ emission in particular, lies within the edge of the dusty component in protoplanetary disks contrasts with previous interferometric studies of young stars. Previous work found the Brγ emission to lie at radii equal to or larger than the dust, implying an origin of the emission in an outflowing wind[7,8]. The differing results for MWC 480 and the two previously observed sources may stem from different luminosities of the objects. MWC 480 is orders of magnitude less luminous than MWC 297 (ref. 15), for which the Brγ emission is more extended than the dust emission[7], and approximately half as luminous at HD 104237 (ref. 27), for which the Brγ and dust emission have comparable size scales[8]. Brγ emission probably traces both accreting and outflowing components in all of these sources to some extent, but the accretion appears to increasingly dominate the emission for lower-luminosity stars.

The distribution of gas at stellocentric radii <0.1 AU in the inner disk of MWC 480 shows that material is present in this protoplanetary disk at radii comparable to the semi-major axes of short-period exoplanets[28]. This finding supports the hypothesis that giant planet migration is halted in resonances with the inner edge of the gaseous disk[4,26]. Future observations of gaseous disk truncation radii in solar-type T Tauri stars will enable direct comparison with exoplanets, which are predominantly found around solar-type stars.

The presence of water vapour in the inner regions of the protoplanetary disk surrounding MWC 480 constrains how water (and other material) is transported in the disk terrestrial zone. Because water beyond the ice-line is in solid form, a concentration gradient tends to push water vapour out of the inner disk; the fact that I observe this inner-disk vapour suggests continual replenishment through the sublimation of inwardly migrating icy bodies[9]. Water vapour in disk terrestrial regions may also enable the *in situ* formation of water-rich planets via adsorption of water onto dust grains and subsequent growth of these water-laden grains into planetesimals[10]. Finally, as one of the dominant gaseous opacity sources in the near-infrared, water vapour can provide an excellent tracer of the temperature and velocity structure of inner-disk gas[25].

Received 25 January; accepted 17 April 2007.

1. Shu, F. H., Adams, F. C. & Lizano, S. Star formation in molecular clouds—Observation and theory. *Annu. Rev. Astron. Astrophys.* **25,** 23–81 (1987).
2. Safronov, V. S. *Evolution of the Protoplanetary Cloud and Formation of the Earth and the Planets* [in Russian] NASA TTF-677 (Nauka, Moscow, 1969).
3. Raymond, S. N., Quinn, T. & Lunine, J. I. Making other Earths: dynamical simulations of terrestrial planet formation and water delivery. *Icarus* **168,** 1–17 (2004).
4. Lin, D. N. C., Bodenheimer, P. & Richardson, D. C. Orbital migration of the planetary companion of 51 Pegasi to its present location. *Nature* **380,** 606–607 (1996).
5. Königl, A. Disk accretion onto magnetic T Tauri stars. *Astrophys. J.* **370,** L39–L43 (1991).
6. Millan-Gabet, R. *et al.* in *Protostars and Planets V* (eds Reipurth, B., Jewitt, D. & Keil, K.) 539–554 (Univ. Arizona Press, Tucson, 2007).
7. Malbet, F. *et al.* Disk and wind interaction in the young stellar object MWC 297 spatially resolved with AMBER/VLTI. *Astron. Astrophys.* **464,** 43–53 (2007).
8. Tatulli, E. *et al.* Constraining the wind launching region in Herbig Ae stars: AMBER/VLTI spectroscopy of HD 104237. *Astron. Astrophys.* **464,** 55–58 (2007).
9. Ciesla, F. J. & Cuzzi, J. N. The evolution of the water distribution in a viscous protoplanetary disk. *Icarus* **181,** 178–204 (2006).
10. Drake, M. J. Origin of water in the terrestrial planets. *Meteorit. Planet. Sci.* **40,** 519–527 (2005).
11. Eisner, J. A. *et al.* Stellar and molecular radii of a Mira star: first observations with the Keck Interferometer grism. *Astrophys. J.* **654,** L77–L80 (2007).
12. Colavita, M. M. & Wizinowich, P. L. in *Interferometry for Optical Astronomy II* (ed. Traub, W. A.) *Proc. SPIE* **4838,** 79–88 (2003).
13. Colavita, M. *et al.* Observations of DG Tauri with the Keck Interferometer. *Astrophys. J.* **592,** L83–L86 (2003).
14. Mannings, V., Koerner, D. W. & Sargent, A. I. A rotating disk of gas and dust around a young counterpart to β Pictoris. *Nature* **388,** 555–557 (1997).
15. Eisner, J. A., Lane, B. F., Hillenbrand, L., Akeson, R. & Sargent, A. Resolved inner disks around Herbig Ae/Be stars. *Astrophys. J.* **613,** 1049–1071 (2004).
16. Monnier, J. D. *et al.* Few skewed disks found in first closure-phase survey of Herbig Ae/Be stars. *Astrophys. J.* **647,** 444–463 (2006).
17. Eisner, J. A., Chiang, E. I., Lane, B. F. & Akeson, R. L. Spectrally dispersed K-band interferometric observations of Herbig Ae/Be sources: inner disk temperature profiles. *Astrophys. J.* **657,** 347–358 (2007).
18. Osterbrock, D. E. *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei* Ch. 4 (Univ. Science Books, California, 1989).
19. Isella, A. & Natta, A. The shape of the inner rim in proto-planetary disks. *Astron. Astrophys.* **438,** 899–907 (2005).
20. Ludwig, C. B. Measurements of the curves-of-growth of hot water vapor. *Appl. Opt.* **10,** 1057–1073 (1971).
21. Pollack, J. B. *et al.* Composition and radiative properties of grains in molecular clouds and accretion disks. *Astrophys. J.* **421,** 615–639 (1994).
22. Najita, J., Carr, J. S. & Tokunaga, A. T. High-resolution spectroscopy of BR gamma emission in young stellar objects. *Astrophys. J.* **456,** 292–299 (1996).
23. Calvet, N. & Gullbring, E. The structure and emission of the accretion shock in T Tauri stars. *Astrophys. J.* **509,** 802–818 (1998).
24. Muzerolle, J., D'Alessio, P., Calvet, N. & Hartmann, L. Magnetospheres and disk accretion in Herbig Ae/Be stars. *Astrophys. J.* **617,** 406–417 (2004).
25. Carr, J. S., Tokunaga, A. T. & Najita, J. Hot H_2O emission and evidence for turbulence in the disk of a young star. *Astrophys. J.* **603,** 213–220 (2004).
26. Najita, J. R., Carr, J. S., Glassgold, A. E. & Valenti, J. A. in *Protostars and Planets V* (eds Reipurth, B., Jewitt, D. & Keil, K.) 507–522 (Univ. Arizona Press, Tucson, 2007).
27. van den Ancker, M. E., de Winter, D., Tjin, A. & Djie, H. R. E. HIPPARCOS photometry of Herbig Ae/Be stars. *Astron. Astrophys.* **330,** 145–154 (1998).
28. Marcy, G. *et al.* Observed properties of exoplanets: masses, orbits, and metallicities. *Progr. Theor. Phys.* **158** (Suppl.), 24–42 (2005).
29. Dullemond, C. P., Dominik, C. & Natta, A. Passive irradiated circumstellar disks with an inner hole. *Astrophys. J.* **560,** 957–969 (2001).
30. Shu, F. *et al.* Magnetocentrifugally driven flows from young stars and disks. 1: A generalized model. *Astrophys. J.* **429,** 781–796 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Data presented herein were obtained at the W. M. Keck Observatory from telescope time allocated to the National Aeronautics and Space Administration through the agency's scientific partnership with the California Institute of Technology and the University of California. The Observatory was made possible by the generous financial support of the W. M. Keck Foundation. I thank the entire Keck Interferometer team for their invaluable contributions to these observations. I also acknowledge input into this work (and this manuscript) from R. Akeson, E. Chiang, A. Glassgold, J. Graham, J. Najita and R. White. I am supported by a Miller Research Fellowship.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.A.E. (jae@astro.berkeley.edu).

LETTERS

Quantum oscillations and the Fermi surface in an underdoped high-T_c superconductor

Nicolas Doiron-Leyraud[1], Cyril Proust[2], David LeBoeuf[1], Julien Levallois[2], Jean-Baptiste Bonnemaison[1], Ruixing Liang[3,4], D. A. Bonn[3,4], W. N. Hardy[3,4] & Louis Taillefer[1,4]

Despite twenty years of research, the phase diagram of high-transition-temperature superconductors remains enigmatic[1,2]. A central issue is the origin of the differences in the physical properties of these copper oxides doped to opposite sides of the superconducting region. In the overdoped regime, the material behaves as a reasonably conventional metal, with a large Fermi surface[3,4]. The underdoped regime, however, is highly anomalous and appears to have no coherent Fermi surface, but only disconnected 'Fermi arcs'[5,6]. The fundamental question, then, is whether underdoped copper oxides have a Fermi surface, and if so, whether it is topologically different from that seen in the overdoped regime. Here we report the observation of quantum oscillations in the electrical resistance of the oxygen-ordered copper oxide $YBa_2Cu_3O_{6.5}$, establishing the existence of a well-defined Fermi surface in the ground state of underdoped copper oxides, once superconductivity is suppressed by a magnetic field. The low oscillation frequency reveals a Fermi surface made of small pockets, in contrast to the large cylinder characteristic of the overdoped regime. Two possible interpretations are discussed: either a small pocket is part of the band structure specific to $YBa_2Cu_3O_{6.5}$ or small pockets arise from a topological change at a critical point in the phase diagram. Our understanding of high-transition-temperature (high-T_c) superconductors will depend critically on which of these two interpretations proves to be correct.

The electrical resistance of two samples of ortho-II ordered $YBa_2Cu_3O_{6.5}$ was measured in a magnetic field of up to 62 T applied normal to the CuO_2 planes ($B||c$). (Sample characteristics and details of the measurements are given in the Methods section.) With a T_c of 57.5 K, these samples have a hole doping per planar copper atom of $p = 0.10$, that is, they are well into the underdoped region of the phase diagram (see Fig. 1a). Angle-resolved photoemission spectroscopy (ARPES) data for underdoped $Na_{2-x}Ca_xCu_2O_2Cl_2$ (Na-CCOC) at precisely the same doping (reproduced in Fig. 1b from ref. 6) shows most of the spectral intensity to be concentrated in a small region near the nodal position ($\pi/2$, $\pi/2$), suggesting a Fermi surface broken up into disconnected arcs, while ARPES studies on overdoped $Tl_2Ba_2CuO_{6+\delta}$ (Tl-2201) at $p = 0.25$ reveal a large, continuous cylinder (reproduced in Fig. 1c from ref. 4).

The Hall resistance R_{xy} as a function of magnetic field is displayed in Fig. 2 for sample A, and in Supplementary Fig. 1 for sample B, where oscillations are clearly seen above the resistive superconducting transition. Note that a vortex liquid phase is believed to extend well above the irreversibility field, beyond our highest field of 62 T, which may explain why R_{xy} is negative at these low temperatures, as opposed to positive at temperatures above T_c. Nevertheless, quantum oscillations are known to exhibit the very same diagnostic characteristics of frequency and mass in the vortex state as in the field-induced normal state above the upper critical field $H_{c2}(0)$ (for example, ref. 7). They are caused by the passage of quantized Landau levels across the Fermi level as the applied magnetic field is varied, and as such they are considered the most robust and direct signature of a coherent Fermi surface. The inset of Fig. 2 shows the 2 K isotherm and a smooth background curve. We extract the oscillatory component, plotted in Fig. 3a as a function

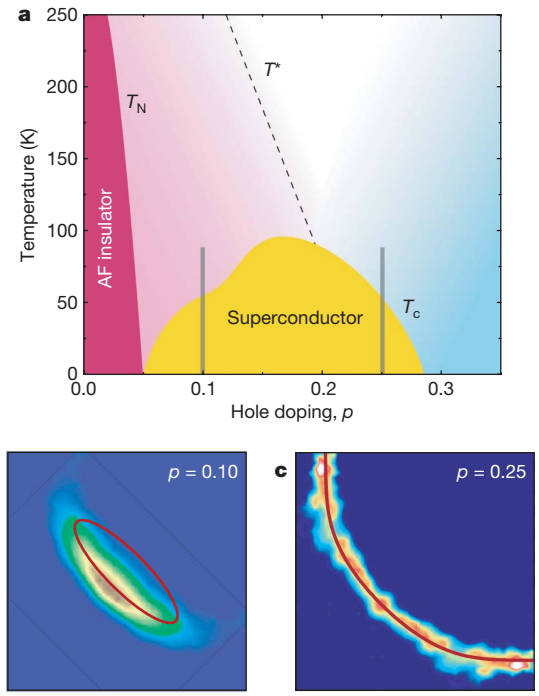


Figure 1 | Phase diagram of high-temperature superconductors. **a**, Schematic doping dependence of the antiferromagnetic (T_N) and superconducting (T_c) transition temperatures and the pseudogap crossover temperature T^* in YBCO. The vertical lines at $p = 0.1$ and $p = 0.25$ mark the positions of copper oxide materials discussed in the text: ortho-II ordered $YBa_2Cu_3O_{6.5}$ and Na-CCOC, located well into the underdoped region, and Tl-2201, well into the overdoped region, respectively. **b**, **c**, Distribution of ARPES spectral intensity in one quadrant of the Brillouin zone, measured (**b**), on Na-CCOC at $p = 0.1$, and (**c**), on Tl-2201 at $p = 0.25$ (reproduced from ref. 6 and ref. 4, with permissions from K. M. Shen and A. Damascelli, respectively). These respectively reveal a truncated Fermi surface made of 'Fermi arcs' at $p = 0.10$, and a large, roughly cylindrical and continuous Fermi surface at $p = 0.25$. The red ellipse in **b** encloses an area A_k that corresponds to the frequency F of quantum oscillations measured in YBCO.

[1]Département de physique and RQMP, Université de Sherbrooke, Sherbrooke, Canada J1K 2R1. [2]Laboratoire National des Champs Magnétiques Pulsés (LNCMP), UMR CNRS-UPS-INSA 5147, Toulouse 31400, France. [3]Department of Physics and Astronomy, University of British Columbia, Vancouver, Canada V6T 1Z4. [4]Canadian Institute for Advanced Research, Toronto, Canada M5G 1Z8.

of inverse field, by subtracting the monotonic background (shown for all temperatures in Supplementary Fig. 2). This shows that the oscillations are periodic in $1/B$, as is expected of oscillations that arise from Landau quantization. A Fourier transform yields the power spectrum, displayed in Fig. 3b, which consists of a single frequency, $F = (530 \pm 20)$ T. In Fig. 3c, we plot the amplitude of the oscillations as a function of temperature, from which we deduce a carrier mass $m^* = (1.9 \pm 0.1)m_0$, where m_0 is the bare electron mass. Within error bars, both F and m^* are the same in sample B, for which the current J is parallel to the b axis (see Supplementary Fig. 1). Oscillations of the same frequency are also observed in R_{xx} (in both samples), albeit with a smaller amplitude. We note that while at 7.5 K the oscillations are still perceptible, they are absent at 11 K, as expected from thermally damped quantum oscillations (see Supplementary Fig. 5).

While quantum oscillations in $YBa_2Cu_3O_{6+y}$ (YBCO) have been the subject of a number of earlier studies[8–10], the data reported so far do not exhibit clear oscillations as a function of $1/B$ and, as such, have not been accepted as convincing evidence for a Fermi surface[11]. Furthermore, we note that all previous work was done on oriented powder samples as opposed to the high-quality single crystals used in the present study.

Quantum oscillations are a direct measure of the Fermi surface area via the Onsager relation: $F = (\Phi_0/2\pi^2)A_k$, where $\Phi_0 = (2.07 \times 10^{-15})$ T m^2 is the flux quantum, and A_k is the cross-sectional area of the Fermi surface normal to the applied field. A frequency of 530 T implies a Fermi surface pocket that encloses a k-space area (in the a–b plane) of $A_k = 5.1$ nm^{-2}, that is, 1.9% of the Brillouin zone (of area $4\pi^2/ab$). This is only 3% of the area of the Fermi surface cylinder measured in Tl-2201 (see Fig. 1c), whose radius is $k_F \approx 7$ nm^{-1}. In the remainder, we examine two scenarios to explain the dramatic difference between the small Fermi surface revealed by the low frequency of quantum oscillations reported here for $YBa_2Cu_3O_{6.5}$ and the large cylindrical surface observed in overdoped Tl-2201. The first scenario assumes that the particular band structure of $YBa_2Cu_3O_{6.5}$ is different and supports a small Fermi surface sheet. In the second, the electronic structure of overdoped copper oxides undergoes a transformation as the doping p is reduced below a value p_c associated with a critical point.

Band structure calculations for stoichiometric YBCO ($y = 1.0$), which is slightly overdoped (with $p = 0.2$), show a Fermi surface consisting of four sheets[12,13], as reproduced in Fig. 4a: two large cylinders derived from the CuO_2 bi-layer, one open surface coming from the CuO chains, and a small cylinder associated with both chain and plane states. The latter sheet, for example, could account for the low frequency reported here. ARPES studies on YBCO near optimal doping[14,15] appear to be in broad agreement with this electronic structure. However, recent band structure calculations[16] performed specifically for $YBa_2Cu_3O_{6.5}$, which take into account the unit cell doubling caused by the ortho-II order, give a Fermi surface where the small cylinder is absent, as shown in Fig. 4b. This leaves no obvious candidate Fermi surface sheet for the small orbit reported here.

The fact that the same oscillations are observed for currents along a and b suggests that they are not associated with open orbits in the chain-derived Fermi surface sheet. In YBCO, the CuO chains along the b axis are an additional channel of conduction, responsible for an anisotropy in the zero-field resistivity $\rho(T)$ of the normal state (above

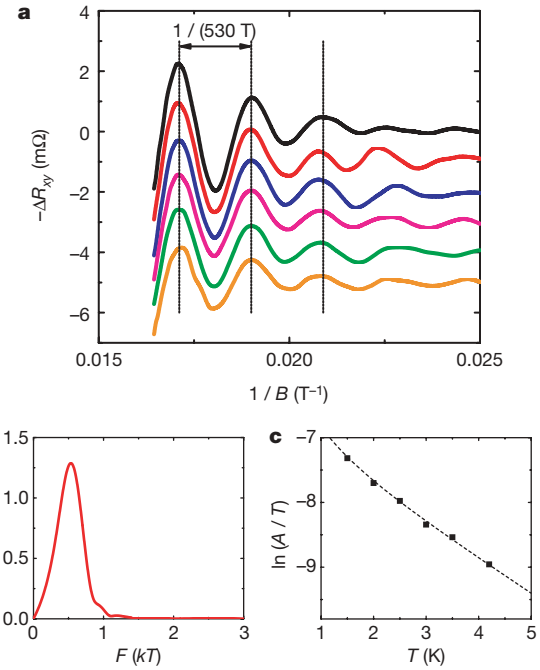


Figure 3 | Quantum oscillations in YBCO. **a**, Oscillatory part of the Hall resistance, obtained by subtracting the monotonic background (shown in the inset of Fig. 2 for $T = 2$ K), as a function of inverse magnetic field, $1/B$. The background at each temperature is given in Supplementary Fig. 2. **b**, Power spectrum (Fourier transform) of the oscillatory part for the $T = 2$ K isotherm, revealing a single frequency at $F = (530 \pm 20)$ T, which corresponds to a k-space area $A_k = 5.1$ nm^{-2}, from the Onsager relation $F = (\Phi_0/2\pi^2)A_k$. Note that the uncertainty of 4% on F is not given by the width of the peak (a consequence of the small number of oscillations), but by the accuracy with which the position of successive maxima in **a** can be determined. **c**, Temperature dependence of the oscillation amplitude A, plotted as $\ln(A/T)$ versus T. The fit is to the standard Lifshitz–Kosevich formula, whereby $A/T = [\sinh(am^*T/B)]^{-1}$, which yields a cyclotron mass $m^* = (1.9 \pm 0.1)m_0$, where m_0 is the free electron mass.

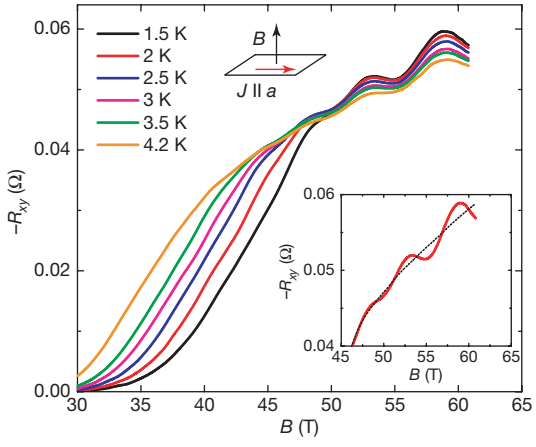


Figure 2 | Hall resistance of $YBa_2Cu_3O_{6.5}$. R_{xy} as a function of magnetic field B, for sample A, at different temperatures between 1.5 and 4.2 K. The field is applied normal to the CuO_2 planes ($B \| c$) and the current is along the a axis of the orthorhombic crystal structure ($J \| a$). The inset shows a zoom on the data at $T = 2$ K, with a fitted monotonic background (dashed line).

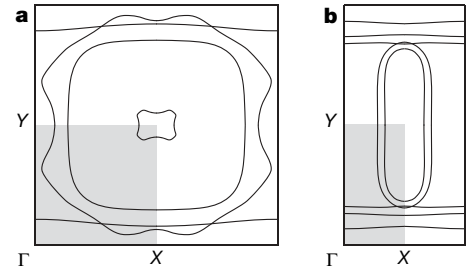


Figure 4 | Fermi surface of YBCO from band structure calculations. **a**, Fermi surface of $YBa_2Cu_3O_7$ in the $k_z = 0$ plane (from ref. 13, with permission from O. K. Andersen), showing the four bands discussed in the main text. **b**, Fermi surface of ortho-II ordered $YBa_2Cu_3O_{6.5}$ in the $k_z = 0$ plane (from ref. 16, with permission from T. M. Rice). In both **a** and **b** the grey shading indicates one quadrant of the first Brillouin zone.

T_c). In Supplementary Fig. 3, we plot the anisotropy ratio ρ_a/ρ_b as a function of temperature, and the chain resistivity versus T^2, obtained from samples A and B. Although the anisotropy is about 2.2 at 300 K, it eventually disappears as $T \rightarrow T_c$, with $\rho_a/\rho_b \approx 1.0$ for $T < 80$ K, owing to the localization of carriers in the quasi-one-dimensional chains. We thus conclude that chains make a negligible contribution to electrical transport at the low temperatures where quantum oscillations are observed. The oscillations must therefore come from the electronic structure of the CuO_2 planes.

The apparent lack of agreement between our experimental data and the calculated band structure of ortho-II $YBa_2Cu_3O_{6.5}$ calls for a careful re-examination of the band structure of YBCO near $p = 0.1$, both with and without ortho-II order. Let us then consider the possibility that electron correlations and/or the onset of an ordered phase with broken symmetry bring about a radical change in the Fermi surface from that predicted by local density approximation (LDA) calculations.

Much like in chromium where a spin-density-wave order causes the Fermi surface to reconstruct, a transformation in the electronic structure of YBCO could take place as a function of doping, resulting in small Fermi surface pockets. Such a scenario was invoked to explain the electron-doped copper oxides[17]. Comparison of A_k in YBCO with ARPES measurements[6] on Na-CCOC at the same doping ($p = 0.1$) suggests that the small Fermi surface pockets observed here are centred at nodal positions, that is to say at the crossing of the original large Fermi surface cylinder with the diagonal from (0, 0) to (π, π), that is near ($\pi/2$, $\pi/2$). In Fig. 1b, we show how an elliptical Fermi surface pocket with an area of 5 nm^{-2} located at this position would indeed match the ARPES spectral weight. Extending this to the four nodal positions could be a solution to the intriguing problem of the 'Fermi arcs', revealing them to be elongated Fermi pockets. Let us examine some implications of this four-pocket scenario.

Assuming that the Fermi surface is strictly two-dimensional, that is, neglecting any c axis dispersion, then the carrier density n per plane per area is given by Luttinger's theorem as $n = 2A_k/(2\pi)^2 = F/\Phi_0$ for each pocket. For four pockets, this gives $n = 1.02\ nm^{-2} = 0.152 \pm 0.006$ carriers per planar Cu atom ($ab = (0.38227 \times 0.38872)\ nm^2$ is the area per Cu atom). Thus we find that $n/p = 1.5$. Taken literally, this would imply that the addition of one hole doped into the CuO_2 plane does not result in identically one extra charge carrier. Or, in other words, the Luttinger sum rule that relates the number of mobile carriers to the Fermi surface volume (or area in two dimensions) appears to be violated. This intriguing possibility depends on a number of assumptions that will require careful examination.

First, we assume that the Onsager relation applies in copper oxides exactly as it does in normal metals, that is, it is the frequency of oscillations in inverse applied magnetic field, rather than some other internal field, which is related to the Fermi surface area via $F = (\Phi_0/2\pi^2)A_k$. Second, we ignore the possibility that electron Fermi surface pockets may also exist, although it is not clear why they were not seen here. Finally, we neglect any dispersion along the c axis. However, to recover the Luttinger sum rule solely on this ground, a rather strong dispersion is necessary, something not predicted by band calculations nor evidenced by an additional quantum oscillation frequency.

For comparison, the carrier density in overdoped Tl-2201 at $p = 0.25$ is $n = (1.26 \pm 0.04)$ hole(s) per Cu atom, as obtained by the Hall effect[18], angular magnetoresistance oscillations (AMRO)[3] and ARPES[4]. In Fig. 1c, this is directly given by the measured Fermi surface area. This is precisely the value expected in the large-Fermi surface scenario of band theory, where $n = 1 + p$. Extending this down to $p = 0.1$ would predict that n should be 1.1 in $YBa_2Cu_3O_{6.5}$. The fact that we get $n = 0.15$ when we assume that the Fermi surface consists of no other sheet than the one manifest in the single frequency observed in our experiment means one of two things: either it is incorrect to extend the $n = 1 + p$ relation down into the underdoped regime, or large parts of the Fermi surface of YBCO were not seen in our experiment (possibly because of a much larger m^*).

In other words, in the four-nodal pocket scenario a transition between a large-Fermi surface metal (in the overdoped regime) and a small-Fermi surface metal (in the underdoped regime) occurs at a critical doping somewhere between $p = 0.1$ and $p = 0.25$. The associated drop in carrier density n by one order of magnitude would be expected to show up in most physical properties, for example the superfluid density n_S and the normal-state electronic specific heat coefficient γ_N. Assuming the dynamical mass $m_d^* \approx 2m_0$, close to the thermal mass m^* deduced from quantum oscillations, we get $n_S = m_d^* c^2/(4\pi e^2 \lambda^2) \approx 0.12$ carriers per planar Cu atom, from the experimentally measured in-plane penetration depth of ortho-II ordered YBCO: $\lambda_a = (200 \pm 20)$ nm (ref. 19). This low value of n_S is much closer to the carrier density obtained in the four-nodal-pocket scenario ($n = 0.15$) than in the large-Fermi-surface scenario ($n = 1 + p = 1.1$). Note, however, that a number of factors need to be considered carefully to make such a quantitative comparison compelling, including the real dynamical mass and the possibility of strong phase fluctuations. The loss in density of states that would follow a Fermi surface transformation into small pockets could also explain the large and sudden drop in γ_N that has been observed in various copper oxides as the doping is reduced below $p \approx 0.19$ (ref. 20).

A number of theories predict a Fermi surface made of four small pockets at nodal positions in the underdoped regime, going over to a large Fermi surface when p exceeds a critical value p_c. Some of these are analogous to the usual spin-density-wave scenario[17] in the sense that they invoke the onset of an ordered phase with broken symmetry below p_c (refs 21–23), while others do not require any broken symmetry[24–26].

In summary, our unambiguous observation of quantum oscillations in underdoped YBCO proves the existence of a Fermi surface. The small size of the Fermi surface pocket associated with the low oscillation frequency suggests two very different scenarios for the non-superconducting ground state of underdoped copper oxides. The first is a multi-band scenario in which the ground state is described by the LDA band structure. In the second scenario, the pseudogap phase that lies to the left of T^* in the phase diagram is a highly correlated electronic fluid with a Fermi surface made of small pockets at nodal positions, separated from the Fermi liquid of the overdoped regime by a critical point near optimal doping.

METHODS SUMMARY

Samples. The samples used are fully detwinned crystals of $YBa_2Cu_3O_{6+y}$ grown in non-reactive $BaZrO_3$ crucibles from high-purity starting materials (see ref. 27). The oxygen content was set at $y = 0.51$ and the dopant oxygen atoms were made to order into an ortho-II superstructure of alternating full and empty CuO_y chains, yielding a superconducting transition temperature $T_c = 57.5$ K. The samples are uncut, unpolished thin platelets, whose transport properties are measured via gold evaporated contacts (contact resistance $<1\ \Omega$), in a six-contact geometry. Sample A (current along a axis) and sample B (current along b axis) have dimensions ($40 \times 520 \times 720$) μm^3 (thickness × length × width) and ($65 \times 810 \times 1{,}030$) μm^3, respectively.

Estimates of hole doping. The hole doping p in YBCO is determined from a relationship between T_c and the c axis lattice constant (see ref. 28). For our samples, the measured $T_c = 57.5$ K implies $p = 0.099$ and the measured $c = (1.17441 \pm 0.00005)$ nm gives $p = 0.098 \pm 0.001$ (ref. 28).

Resistance measurements. Longitudinal (R_{xx}) and transverse (R_{xy}) resistances are obtained from the voltage difference measured diagonally on either side of the sample width, for a field parallel (up) and anti-parallel (down) to the c axis: $R_{xx} \equiv (V_{up} + V_{down})/2I_x$ and $R_{xy} \equiv (V_{up} - V_{down})/2I_x$. A current excitation of 5 mA at 40 kHz was used. The voltage (and a reference signal) was digitized using a high-speed digitizer and post-analysed to perform the phase comparison. The measurements were performed at the LNCMP in Toulouse, in a pulsed resistive magnet of up to 62 T (ref. 29). Data for the rise (26 ms) and fall (110 ms) of the field pulse were in perfect agreement, thus excluding any heating due to eddy current.

Received 3 April; accepted 18 April 2007.

1. Orenstein, J. & Millis, A. J. Advances in the physics of high-temperature superconductivity. *Science* **288,** 468–474 (2000).
2. Norman, M. R., Pines, D. & Kallin, C. The pseudogap: friend or foe of high T_c? *Adv. Phys.* **54,** 715–733 (2005).
3. Hussey, N. E. *et al.* Observation of a coherent three-dimensional Fermi surface in a high-transition temperature superconductor. *Nature* **425,** 814–817 (2003).
4. Platé, M. *et al.* Fermi surface and quasiparticle excitations of overdoped $Tl_2Ba_2CuO_{6+\delta}$ by ARPES. *Phys. Rev. Lett.* **95,** 077001 (2005).
5. Norman, M. *et al.* Destruction of the Fermi surface in underdoped high-T_c superconductors. *Nature* **392,** 157–160 (1998).
6. Shen, K. *et al.* Nodal quasiparticles and antinodal charge ordering in $Ca_{2-x}Na_xCuO_2Cl_2$. *Science* **307,** 901–904 (2005).
7. Wosnitza, J. *et al.* Shubnikov-de Haas effect in the superconducting state of an organic superconductor. *Phys. Rev. B* **62,** 11973–11976 (2000).
8. Fowler, C. M. *et al.* de Haas-van Alphen effect and Fermi surface of $YBa_2Cu_3O_{6.97}$. *Phys. Rev. Lett.* **68,** 534–537 (1992).
9. Kido, G., Komorita, K., Katayama-Yoshida, H. & Takahashi, T. de Haas-van Alphen measurement of $YBa_2Cu_3O_7$. *J. Phys. Chem. Solids* **52,** 1465–1470 (1991).
10. Haanappel, E. G. *et al.* The de Haas-van Alphen effect in $YBa_2Cu_3O_{7-\delta}$. *J. Phys. Chem. Solids* **54,** 1261–1267 (1993).
11. Springford, M., Harrison, N., Meeson, P. & Probst, P.-A. Comment on "de Haas-van Alphen effect and Fermi surface of $YBa_2Cu_3O_{6.97}$". *Phys. Rev. Lett.* **69,** 2453 (1992).
12. Pickett, W. E., Cohen, R. E. & Krakauer, H. Precise band structure and Fermi-surface calculation for $YBa_2Cu_3O_7$: Importance of three-dimensional dispersion. *Phys. Rev. B* **42,** 8764–8767 (1990).
13. Andersen, O. K., Liechtenstein, A. I., Jepsen, O. & Paulsen, F. LDA energy bands, low-energy Hamiltonians, t', t'', $t_\perp(\mathbf{k})$, and $J_\perp$. *J. Phys. Chem. Solids* **56,** 1573–1591 (1995).
14. Campuzano, J. C. *et al.* Fermi surfaces of $YBa_2Cu_3O_{6.9}$ as seen by angle-resolved photoemission. *Phys. Rev. Lett.* **64,** 2308–2311 (1990).
15. Nakayama, K. *et al.* Bulk and surface low-energy excitations in $YBa_2Cu_3O_{7-\delta}$ studied by high-resolution angle-resolved photoemission spectroscopy. *Phys. Rev. B* **75,** 014513 (2007).
16. Bascones, E. *et al.* Optical conductivity of ortho-II $YBa_2Cu_3O_{6.5}$. *Phys. Rev. B* **71,** 012505 (2005).
17. Lin, J. & Millis, A. J. Theory of low-temperature Hall effect in electron-doped cuprates. *Phys. Rev. B* **72,** 214506 (2005).
18. Mackenzie, A. P., Julian, S. R., Sinclair, D. C. & Lin, C. T. Normal-state magneto-transport in superconducting $Tl_2Ba_2CuO_{6+y}$ to millikelvin temperatures. *Phys. Rev. B* **53,** 5848–5855 (1996).
19. Pereg-Barnea, T. *et al.* Absolute values of the London penetration depth in $YBa_2Cu_3O_{6+y}$ measured by zero field ESR spectroscopy on Gd doped single crystals. *Phys. Rev. B* **69,** 184513 (2004).
20. Loram, J. W. *et al.* Specific heat evidence on the normal state pseudogap. *J. Phys. Chem. Solids* **59,** 2091–2094 (1998).
21. Chakravarty, S., Laughlin, R. B., Morr, D. K. & Nayak, C. Hidden order in the cuprates. *Phys. Rev. B* **63,** 094503 (2001).
22. Lee, P. A. & Wen, X.-G. Vortex structure in underdoped cuprates. *Phys. Rev. B* **63,** 224517 (2001).
23. Oganesyan, V., Kivelson, S. A. & Fradkin, E. Quantum theory of a nematic Fermi fluid. *Phys. Rev. B* **64,** 195109 (2001).
24. Yang, K.-Y., Rice, T. M. & Zhang, F.-C. Phenomenological theory of the pseudogap state. *Phys. Rev. B* **73,** 174501 (2006).
25. Stanescu, T. D. & Kotliar, G. Fermi arcs and hidden zeros of the Green function in the pseudogap state. *Phys. Rev. B* **74,** 125110 (2006).
26. Kyung, B. *et al.* Pseudogap induced by short-range spin correlations in a doped Mott insulator. *Phys. Rev. B* **73,** 165114 (2006).
27. Liang, R., Bonn, D. A. & Hardy, W. N. Preparation and X-ray characterization of highly ordered ortho-II phase $YBa_2Cu_3O_{6.50}$ single crystals. *Physica C* **336,** 57–62 (2000).
28. Liang, R., Bonn, D. A. & Hardy, W. N. Evaluation of CuO_2 plane hole doping in $YBa_2Cu_3O_{6+x}$ single crystals. *Phys. Rev. B* **73,** 180505 (2006).
29. Portugall, O. *et al.* The LNCMP: a pulsed-field user-facility in Toulouse. *Physica B* **346–347,** 668–672 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. T. Brisson, G. G. Lonzarich, G. L. J. A. Rikken and A.-M. S. Tremblay for discussions, and M. Nardone and A. Audouard for their help with the experiment and analysis. We acknowledge support from the Canadian Institute for Advanced Research and the LNCMP, and funding from NSERC, FQRNT and a Canada Research Chair. Part of this work was supported by the French ANR IceNET and EuroMagNET.

Author Contributions N.D.-L. and C.P. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.P. (proust@lncmp.org) or L.T. (louis.taillefer@physique.usherbrooke.ca).

LETTERS

Visualizing pair formation on the atomic scale in the high-T_c superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$

Kenjiro K. Gomes[1]*, Abhay N. Pasupathy[1]*, Aakash Pushp[1]*, Shimpei Ono[2], Yoichi Ando[2] & Ali Yazdani[1]

Pairing of electrons in conventional superconductors occurs at the superconducting transition temperature T_c, creating an energy gap Δ in the electronic density of states (DOS)[1]. In the high-T_c superconductors, a partial gap in the DOS exists for a range of temperatures above T_c (ref. 2). A key question is whether the gap in the DOS above T_c is associated with pairing, and what determines the temperature at which incoherent pairs form. Here we report the first spatially resolved measurements of gap formation in a high-T_c superconductor, measured on $Bi_2Sr_2CaCu_2O_{8+\delta}$ samples with different T_c values (hole concentration of 0.12 to 0.22) using scanning tunnelling microscopy. Over a wide range of doping from 0.16 to 0.22 we find that pairing gaps nucleate in nanoscale regions above T_c. These regions proliferate as the temperature is lowered, resulting in a spatial distribution of gap sizes in the superconducting state[3–5]. Despite the inhomogeneity, we find that every pairing gap develops locally at a temperature T_p, following the relation $2\Delta/k_BT_p = 7.9 \pm 0.5$. At very low doping ($\leq$0.14), systematic changes in the DOS indicate the presence of another phenomenon[6–9], which is unrelated and perhaps competes with electron pairing. Our observation of nanometre-sized pairing regions provides the missing microscopic basis for understanding recent reports[10–13] of fluctuating superconducting response above T_c in hole-doped high-T_c copper oxide superconductors.

The pseudogap state between T_c and the temperature T^*, below which the gap in the DOS occurs, has been the subject of a wide range of theoretical proposals—from those focused on superconducting pairing correlations without phase coherence[14,15] to those based on some form of competing electronic order or proximity to the Mott state[6–9]. Some experiments have shown evidence of pairing correlations above bulk T_c (refs 10–13), while other experiments have associated the pseudogap with phenomena other than pairing, such as real space electronic organization, which is prominent at low dopings[16–18]. Various tunnelling experiments[19–21] have examined the relationship between the pseudogap and pair formation; however, the spatial variation of DOS and gaps on the nanoscale[3–5] complicates their interpretation and has prevented researchers from reaching a consistent conclusion. To obtain precise answers to these questions, it is necessary to perform systematic atomic-resolution scanning tunnelling microscopy (STM) measurements as a function of both doping and temperature.

We have measured the temperature evolution of the DOS of the high-T_c superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$ using a specially designed variable temperature ultrahigh-vacuum STM. Our instrument allows us to track specific areas of the sample on the atomic scale as a function of temperature. To develop a systematic approach for the analysis of the data at different dopings and temperatures, we first consider the measurements on the most overdoped sample, with $x = 0.22$. At this doping, pseudogap effects have been reported to be either weak or absent[2,6], allowing us to interpret gaps in the DOS as those associated with superconductivity.

Figure 1 shows spectroscopy measurements for an overdoped sample with $T_c = 65$ K (OV65) performed at a specific atomic site over a range of temperatures close to T_c. From such spectra, we determine

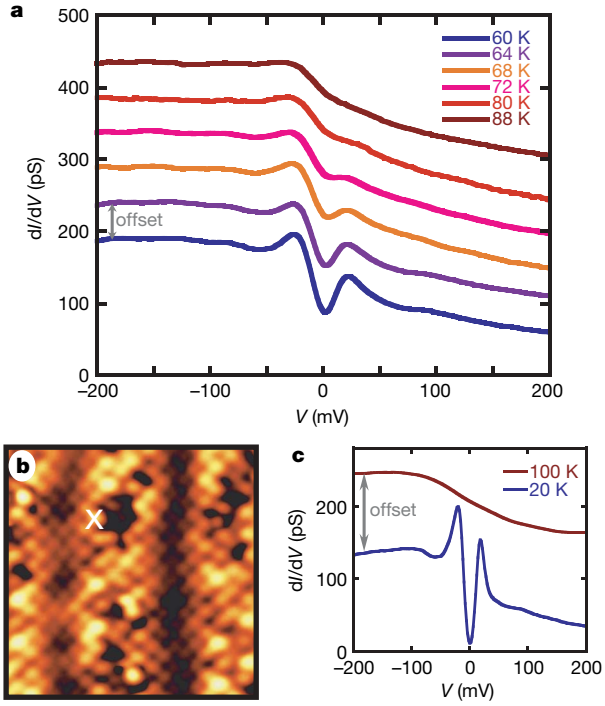


Figure 1 | Spectroscopy at a specific atomic site as a function of temperature (*dI*/*dV* versus sample bias *V*). Samples were cleaved *in situ* at room temperature before being inserted into the cooled microscope stage and measured at various temperatures (20–180 K). **a**, Spectra taken at different temperatures at the same location in an OV65 ($T_c = 65$ K) sample (offset for clarity). Special experimental procedures have been used to track the same location on the sample within 0.1 Å at different temperatures. **b**, Topography of the sample showing the location for spectra in **a**. Between each measurement the STM is stabilized at each temperature for a period of up to 24 h, during which STM topography, such as that shown in **b**, is used to centre the location of spectroscopic measurements shown in **a**. **c**, Typical spectra at 20 and 100 K for the same sample (not the same location). The coherence peaks become more pronounced at lower temperatures, but the position of the peak does not change significantly below 60 K. Analysis of the low-temperature spectra (see Supplementary Information) shows them to be consistent with those of the excitation spectrum of a *d*-wave superconductor uniformly averaged around the Fermi surface with a maximum pairing gap corresponding to $\Delta \approx 24$ mV.

[1]Department of Physics, Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08544, USA. [2]Central Research Institute of Electric Power Industry, Komae, Tokyo 201-8511, Japan.
*These authors contributed equally to this work.

two important quantities. First, we measure the maximum value of the local gap $\Delta \approx 24$ meV. Second, we estimate the temperature $T_p \approx 72$–80 K at which Δ is no longer measurable, using the criterion that at this temperature $dI/dV(V=0) \geq dI/dV$(for all $V>0$). Above T_p the spectra shows a bias-asymmetric background in the DOS that changes little with increasing temperature—indicating that the pairing gap is either absent or no longer relevant at this atomic site. On the basis of this procedure, we find that the data in Fig. 1a can be described by the relation $2\Delta/k_B T_p \approx 7.7$, where k_B is Boltzmann's constant and T_p is the gap closing temperature. While this measurement at a single atomic site is not statistically significant, it establishes the procedure that we extend to large sets of similar measurements.

The evolution of the pairing gap with temperature can be examined statistically using spectroscopic mapping measurements over large areas (~300 Å) of the sample as a function of temperature. Such experiments allow direct visualization on the atomic scale of the development of gaps. In the superconducting state ($T < T_c$), the overdoped OV65 sample shows (Fig. 2a) a distribution of Δ (refs 3 and 4). With increasing temperature, gap maps and related histograms on the OV65 sample display a rapid increase of ungapped regions (Fig. 2a–e). Although the temperatures over which the gaps collapse locally are relatively close to T_c for this sample, these measurements clearly demonstrate that gaps persist locally on the nanoscale over a range of temperatures T_p. These results are reminiscent of recent calculations of the evolution of pairing gaps with temperature in a *d*-wave superconductor with spatially varying pairing correlations[22].

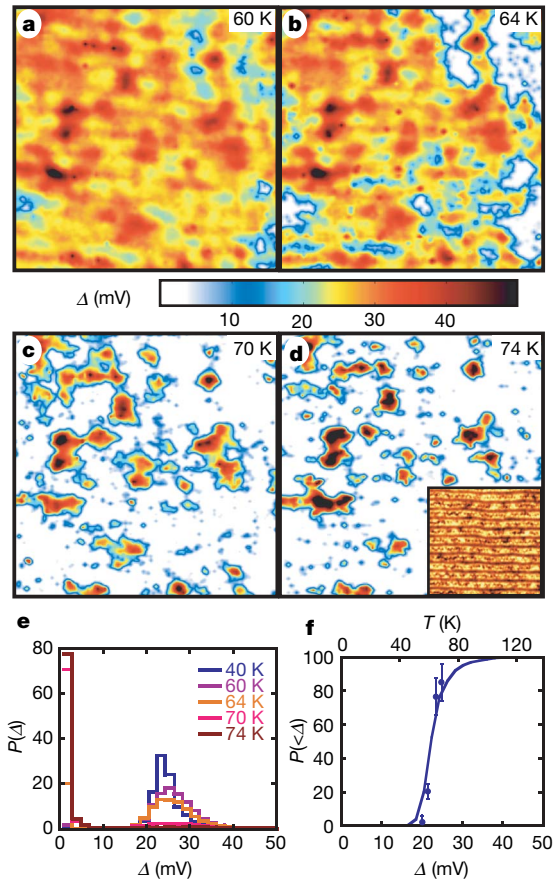


Figure 2 | Gap evolution for overdoped ($T_c = 65$ K) samples. a–d, Gap maps taken on the same 300 Å area of an OV65 sample at different temperatures near T_c. At each temperature and atomic site, the value of the gap can be extracted from local spectroscopic measurements by using the experimental criterion that the local dI/dV has a maximum at $V = +\Delta$, as demonstrated by data in Fig. 1a. The gaps vary spatially on the scale of 1–3 nm. The inset to **d** shows the topography of the area. **e**, The probability of finding a gap of a given size (gap distribution) for the gap maps shown in (**a**–**d**) and an additional gap map taken at 40 K. **f**, The solid line shows the probability $P(<\Delta)$ that the gaps are less than a given Δ (lower *x* axis). This is obtained at a given voltage by summing the lowest temperature histogram of Δ to that voltage. The percentage of ungapped area of the sample (points) is plotted as a function of temperature (shown on the upper *x* axis). The scaling between the two *x* axes is $2\Delta/k_B T = 7.8$. The 1σ error bars were calculated from the finite statistics (arising from the limited area of each map) and the energy and conductance resolution of the spectra used to compose the maps.

We can use the comparison between $P(<\Delta)$ measured at the lowest temperature with the percentage of the ungapped regions measured as a function of temperature to test our local pairing hypothesis for samples at various dopings. The measurements of these two quantities are displayed in Fig. 3e and f, where a single temperature–gap scaling relation $2\Delta/k_B T = 8.0$ has been used to plot data on all samples in this study. From Fig. 3e it is clear that overdoped and optimally doped samples have identical gap–temperature scaling ratios, which, together with the consistency of their low-temperature spectra with a *d*-wave superconducting gap, implies that we can interpret these gaps as those due to pairing. These results clearly show that pairing gaps and the temperature at which they collapse (which can be equal to or larger than T_c) follow a universal local criterion over a wide range of doping. The extracted ratio also shows that the local pairing gap is far more fragile to increases in temperature than are the conventional Bardeen–Cooper–Schrieffer (BCS) superconductors, for which the ratio is in the range of 3.5–5. Surprisingly, the ratio is insensitive to the size of the gap, for gaps ranging from ~15 to 50 mV, indicating that even the smallest gaps are very far from the BCS limit.

Our local pairing hypothesis, however, appears to fail in the underdoped regime (Fig. 3f). Although such a pairing hypothesis has its shortcomings, such as ignoring the possibility of a proximity effect,

we show that the deviation from this picture in underdoped samples is most probably due to the complication of two energy scales in this doping regime. Fig. 4a shows representative spectra measured at temperatures far above T_c that allow us to compare the behaviour of overdoped, optimal and underdoped samples. Once the pairing gaps collapse at high temperatures, the overdoped and optimally doped samples show remarkably similar electron–hole asymmetric spectra. In contrast, the underdoped UD73 sample shows very different, V-shaped spectra with an ill-defined gap, which is insensitive to increasing temperature. Clearly, such V-shaped spectra are related to the strong pseudogap behaviour in underdoped copper oxides[17,18]; however, these spectra and their pseudogap behaviour are qualitatively different from the pseudogaps observed on optimal and overdoped samples above T_c. Such a difference is also seen in ARPES (angle-resolved photoelectron spectroscopy) measurements where the gap closes in overdoped samples but 'fills in' for underdoped samples[23]. A key question is whether all gaps in the underdoped

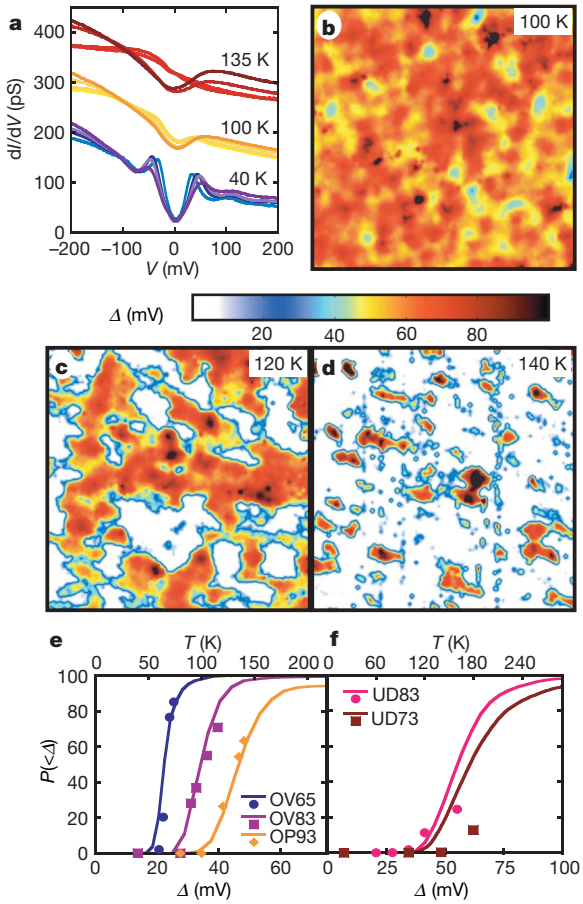


Figure 3 | Gap evolution at different dopings. **a**, Representative spectra taken at different temperatures (offset for clarity) on an OP93 ($T_c = 93$ K) sample. The coherence peaks diminish in intensity on raising the temperature through T_c, but a peak in the spectrum persists at positive bias. **b**–**d**, Gap maps taken over 300 Å areas at different temperatures on the OP93 sample. Although the area over which the maps have been obtained is different for each temperature, the gap distributions are statistically meaningful because of the large areas over which data has been collected. **e**, The solid lines show the integral of the gap distributions versus gap size (lower x axis) for three different doping levels, from overdoped ($T_c = 65$ and 83 K) to optimally doped ($T_c = 93$ K). The points show the percentage of the ungapped area of the samples as a function of temperature (upper x axis). The scaling between the two axes is $2\Delta/k_BT = 8.0$. **f**, The solid lines show the integral of the gap distributions versus gap size (lower axis) for two different doping levels, both underdoped ($T_c = 83$ and 73 K). The points show the percentage of ungapped area of the samples as a function of temperature (upper x axis). The scaling between the two axes is $2\Delta/k_BT = 8.0$.

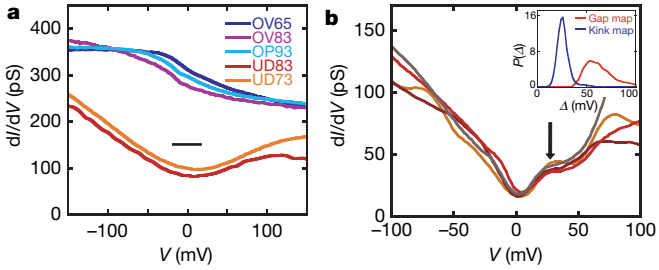


Figure 4 | Pseudogap and pairing gaps in underdoped samples. **a**, Representative spectra taken at specific locations well above T_c for all doping levels (offset for clarity; the bar marks zero conductance for offset spectra). **b**, Spectra taken at 20 K at various locations on the UD73 sample. Within the large gap we see a 'kink' at lower energy, indicated by the arrow. Inset, distributions of the large gap and the 'kink' energy on the UD73 sample at 20 K.

regime as measured by STM can be ascribed to pairing. We find that for $T \ll T_c$, more than 30% of the spectra on the underdoped sample show 'kinks' in the spectra at low bias, indicating the importance of a lower energy scale, as shown in Fig. 4b[3,17,18,24]. The probability of observing such spectra in optimal or overdoped samples is negligible. Spectroscopic mapping with STM can be used to determine the spatial variation and distribution of each energy scale (inset, Fig. 4b). The presence of this additional energy scale indicates that the large gaps seen in underdoped samples cannot be simply associated with pairing. Evidence for two energy scales, one related to pairing and one related to strong pseudogap behaviour in underdoped copper oxides, has been accumulating from photoemission[23–25] and Raman measurements[26]. In contrast, at doping levels beyond optimal doping, results from these experiments can be captured with a single energy scale[23,26]. Momentum resolution (lacking in STM) is important to resolve the two energy scales in underdoped samples. Further work is required to establish firmly whether the 'kink' energy in STM spectra is related to pairing in these samples.

Our ability to visualize the development of gaps and the local pairing hypothesis, which we have established quantitatively on optimal and overdoped samples provides a microscopic picture with which to understand several key aspects of the copper oxide phase diagram. In Fig. 5, we summarize our observations of the spatially inhomogeneous development of the gaps with a colour plot showing the percentage of areas that are gapped as a function of temperature

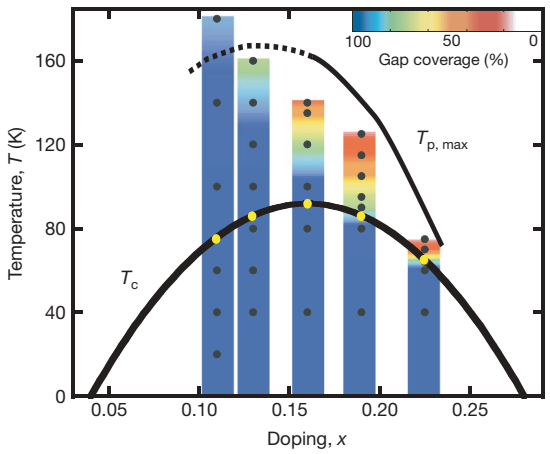


Figure 5 | Schematic phase diagram for $Bi_2Sr_2CaCu_2O_{8+\delta}$. Temperatures and doping levels where large area gap maps were obtained are indicated by points. The colours are the percentage of the sample that is gapped at a given temperature and doping as measured in the gap maps. The $T_{p,max}$ line is the temperature at which <10% of the sample is gapped. The lower solid line denotes the bulk T_c.

at specific dopings. The rising percentage of the gapped regions in various samples with lowering temperature that we probed in our experiments on the atomic scale correlates remarkably well with the onset of the suppression of low-energy excitations probed by other techniques[27]. As we discussed above, for the optimal and overdoped samples a single energy gap can describe all of our findings (Figs 1–3 and additional details in the Supplementary Information), strongly suggesting that the onset of the gap is indeed due to pairing, which occurs locally at T_p. The apparent T^* line is controlled by the largest pairing gaps ($T_{p,max}$) for these samples.

In contrast, for underdoped samples our data supports the notion of two energy scales. The T^* line is controlled by the larger of the two scales, which appears to be unrelated to pairing. Our data suggests that the $T_{p,max}$ line could be well below T^* for low doping. This possibility is also supported by measurements of fluctuating superconductivity, which have shown that the onset temperatures for these fluctuations are well below T^* for underdoped samples[10–13]. A comparison of our data to those from Nernst and magnetization measurements on similar $Bi_2Sr_2CaCu_2O_{8+\delta}$ samples[13] shows that the macroscopic signature of the fluctuating superconducting state appears when ~50% of the sample develops a pairing gap. Given that an adequate amount of pairing has to develop in the samples for the vortex response to be well defined, our measurements provide the missing microscopic basis for the onset of the vortex response. Our observations of local pairing over a range of temperatures, as well as the Nernst and magnetization measurements, show that T_c in the $Bi_2Sr_2CaCu_2O_{8+\delta}$ system marks the onset of phase coherence and not the formation of pairs as in BCS superconductors.

Finally, the local pairing criterion extracted from a large number of measurements on $Bi_2Sr_2CaCu_2O_{8+\delta}$ at various dopings and temperatures has important implications for the mechanism of pairing in copper oxides. In conventional superconductors, the strength of the coupling of electrons to phonons determines both Δ and T_c, with stronger coupling resulting in an increase of both these quantities. However, in the strong-coupling limit, the ratio $2\Delta/k_BT_c$ is dependent on Δ and increases from the universal BCS ratio of 3.5 (ref. 1). The extension of BCS theory based on the Eliashberg equations captures this behaviour for conventional superconductors[28]. Our observation of the insensitivity of the large ratio $2\Delta/k_BT_c = 7.9 \pm 0.5$ to the size of local Δ values (from 15–50 mV)—the local disorder—as well as the doping is hence fundamentally different from the expectations[29] from a electron–boson pairing mechanism based on an Eliashberg-type theory. A successful theory of copper oxides would have to explain not only how pairing correlations can nucleate in isolated nanoscale regions at high temperatures but also the robustness of the local pairing criterion reported here.

Received 23 January; accepted 19 April 2007.

1. Tinkham, M. *Introduction to Superconductivity* (McGraw-Hill, New York, 1975).
2. Timusk, T. & Statt, B. The pseudogap in high-temperature superconductors: an experimental survey. *Rep. Prog. Phys.* **62,** 61–122 (1999).
3. Howald, C., Fournier, P. & Kapitulnik, A. Inherent inhomogeneities in tunnelling spectra of $Bi_2Sr_2CaCu_2O_{8-x}$ crystals in the superconducting state. *Phys. Rev. B* **64,** 100504(R) (2001).
4. Pan, S. H. *et al.* Microscopic electronic inhomogeneity in the high-T_c superconductor $Bi_2Sr_2CaCu_2O_{8+x}$. *Nature* **413,** 282–285 (2001).
5. McElroy, K. *et al.* Atomic-scale sources and mechanism of nanoscale electronic disorder in $Bi_2Sr_2CaCu_2O_{8+\delta}$. *Science* **309,** 1048–1052 (2005).
6. Tallon, J. L. & Loram, J. W. The doping dependence of T*—what is the real high-T_c phase diagram? *Physica C* **349,** 53–68 (2001).
7. Kivelson, S. A. *et al.* How to detect fluctuating stripes in the high-temperature superconductors. *Rev. Mod. Phys.* **75,** 1201–1241 (2003).
8. Norman, M. R., Pines, D. & Kallin, C. The pseudogap: friend or foe of high T_c? *Adv. Phys.* **54,** 715–733 (2005).
9. Lee, P. A., Nagaosa, N. & Wen, X.-G. Doping a Mott insulator: physics of high-temperature superconductivity. *Rev. Mod. Phys.* **78,** 17–86 (2006).
10. Corson, J., Mallozzi, R., Orenstein, J., Eckstein, J. N. & Bozovic, I. Vanishing of phase coherence in underdoped $Bi_2Sr_2CaCu_2O_{8+\delta}$. *Nature* **398,** 221–223 (1999).
11. Xu, Z. A., Ong, N. P., Wang, Y., Kakeshita, T. & Uchida, S. Vortex-like excitations and the onset of superconducting phase fluctuation in underdoped $La_{2-x}Sr_xCuO_4$. *Nature* **406,** 486–488 (2000).
12. Wang, Y. *et al.* Field-enhanced diamagnetism in the pseudogap state of the cuprate $Bi_2Sr_2CaCu_2O_{8+\delta}$ superconductor in an intense magnetic field. *Phys. Rev. Lett.* **95,** 247002 (2005).
13. Wang, Y., Li, L. & Ong, N. P. Nernst effect in high-T_c superconductors. *Phys. Rev. B* **73,** 024510 (2006).
14. Emery, V. J. & Kivelson, S. A. Importance of phase fluctuations in superconductors with small superfluid density. *Nature* **374,** 434–437 (1995).
15. Randeria, M. in *Proc. Int. School of Physics 'Enrico Fermi' on Conventional and High Temperature Superconductors* (eds Iadonisi, G., Schrieffer, J. R. & Chiafalo, M. L.) 53–75 (IOS Press, Amsterdam, 1998).
16. Vershinin, M. *et al.* Local ordering in the pseudogap state of the high-T_c superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$. *Science* **303,** 1995–1998 (2004).
17. Hanaguri, T. *et al.* A 'checkerboard' electronic crystal state in lightly hole-doped $Ca_{2-x}Na_xCuO_2Cl_2$. *Nature* **430,** 1001–1005 (2004).
18. McElroy, K. *et al.* Coincidence of checkerboard charge order and antinodal state decoherence in strongly underdoped superconducting $Bi_2Sr_2CaCu_2O_{8+\delta}$. *Phys. Rev. Lett.* **94,** 197005 (2005).
19. Renner, Ch, Revaz, B., Genoud, J.-Y., Kadowaki, K. & Fischer, Ø. Pseudogap precursor of the superconducting gap in under- and overdoped $Bi_2Sr_2CaCu_2O_{8+\delta}$. *Phys. Rev. Lett.* **80,** 149–152 (1998).
20. Kugler, M., Fischer, Ø., Renner, Ch, Ono, S. & Ando, Y. Scanning tunneling spectroscopy of $Bi_2Sr_2CuO_{6+\delta}$: new evidence for the common origin of the pseudogap and superconductivity. *Phys. Rev. Lett.* **86,** 4911–4914 (2001).
21. Deutscher, G. Coherence and single particle excitations in high temperature superconductors. *Nature* **397,** 410–412 (1999).
22. Andersen, B. M., Melikyan, A., Nunner, T. S. & Hirschfeld, P. J. Thermodynamic transitions in inhomogeneous d-wave superconductors. *Phys. Rev. B* **74,** 060501(R) (2006).
23. Norman, M. R. *et al.* Destruction of the Fermi surface in underdoped high-T_c superconductors. *Nature* **397,** 157–160 (1998).
24. Valla, T., Fedorov, A. V., Lee, J., Davis, J. C. & Gu, G. D. The ground state of the pseudogap in cuprate superconductors. *Science* **314,** 1914–1916 (2006).
25. Tanaka, K. *et al.* Distinct Fermi-momentum-dependent energy gaps in deeply underdoped Bi2212. *Science* **314,** 1910–1913 (2006).
26. Le Tacon, M. *et al.* Two energy scales and two distinct quasiparticle dynamics in the superconducting state of underdoped cuprates. *Nature Phys.* **2,** 537–543 (2006).
27. Damascelli, A., Hussain, Z. & Shen, Z.-X. Angled-resolved photoemission studies of cuprate superconductors. *Rev. Mod. Phys.* **75,** 473–541 (2003).
28. Carbotte, J. P. Properties of boson-exchange superconductors. *Rev. Mod. Phys.* **62,** 1027–1157 (1990).
29. Balatsky, A. V. & Zhu, J.-X. Local strong coupling pairing in d-wave supercondcutors with inhomogeneous bosonic modes. *Phys. Rev. B* **74,** 094517 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by Princeton University, the NSF-DMR, and the NSF-MRSEC programme through the Princeton Center for Complex Materials.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.Y. (yazdani@princeton.edu).

LETTERS

Spin-based logic in semiconductors for reconfigurable large-scale circuits

H. Dery[1], P. Dalal[1], Ł. Cywiński[1] & L. J. Sham[1]

Research in semiconductor spintronics aims to extend the scope of conventional electronics by using the spin degree of freedom of an electron in addition to its charge[1]. Significant scientific advances in this area have been reported, such as the development of diluted ferromagnetic semiconductors[2,3], spin injection into semiconductors from ferromagnetic metals[4–8] and discoveries of new physical phenomena involving electron spin[9,10]. Yet no viable means of developing spintronics in semiconductors has been presented. Here we report a theoretical design that is a conceptual step forward—spin accumulation is used as the basis of a semiconductor computer circuit. Although the giant magnetoresistance effect in metals[11,12] has already been commercially exploited, it does not extend to semiconductor/ferromagnet systems, because the effect is too weak for logic operations. We overcome this obstacle by using spin accumulation rather than spin flow[13–15]. The basic element in our design is a logic gate that consists of a semiconductor structure with multiple magnetic contacts; this serves to perform fast and reprogrammable logic operations in a noisy, room-temperature environment. We then introduce a method to interconnect a large number of these gates to form a 'spin computer'. As the shrinking of conventional complementary metal-oxide–semiconductor (CMOS) transistors reaches its intrinsic limit, greater computational capability will mean an increase in both circuit area and power dissipation. Our spin-based approach may provide wide margins for further scaling and also greater computational capability per gate.

At present, spintronics is realized only in all-metallic systems for applications in magnetic field sensing and non-volatile storage[16]. There have been only a few proposals to use logic gates in all-metallic magnetic systems[17–21]. Progress in this field was made possible by the discovery of the giant magnetoresistance effect in heterostructures of ferromagnetic and paramagnetic metal layers[11,12]. The nascent research in semiconductor spintronics emerged with the seminal proposal of a spin-based current modulator[22], and was accelerated by the discovery of surprisingly long spin lifetimes at low temperatures under certain doping conditions[23]. Injection of spin-polarized current in hybrid ferromagnet-semiconductor systems at room temperature is being constantly improved[4–6].

Figure 1a shows the scheme of a universal magnetologic gate (MLG) set for the NAND operation; any binary logic function can be realized using a finite number of these gates. It consists of a semiconductor layer beneath five elongated magnetic contacts with parallel easy magnetization axes. We define fixed opposite directions as the binary states of '1' and '0', as shown in the figure. The operands of the logic operation are the magnetization directions of the contacts labelled by A, B, X and Y. Addressing a specific contact is facilitated by a matrix of current-carrying wires deposited on top of the structure (and/or below) that orient the magnetizations as in magnetic random access memory (MRAM) architectures[24,25]. The output of the gate is given by the current labelled by $I_M(t)$. This current is a transient response caused by rotating the magnetization of the middle contact, M. Below, we show a clear distinction between the current amplitude for '1' and '0' logic outputs. However, both output currents induce magnetic fields that are too small to write the magnetization direction of other contacts (cascading the output from a logic gate to the input of another). Attempting to amplify this current would lead to extreme power consumption when scaling-up the number of gates. Later, we will describe a circuit which avoids this problem. Figure 1b shows a closely related scheme in which a non-magnetic contact lies beneath the mid-section of the MLG. Here, the magnetization of M is pinned and $I_M(t)$ is triggered by applying a voltage signal (labelled by clk2) on the back-gate. We defer the discussion of this case to the end and proceed with the scheme shown in Fig. 1a.

The MLG in Fig. 1a makes direct use of the distinction between the spin accumulation densities inside the semiconductor for the case of parallel or antiparallel alignment of two contacts at each side of the

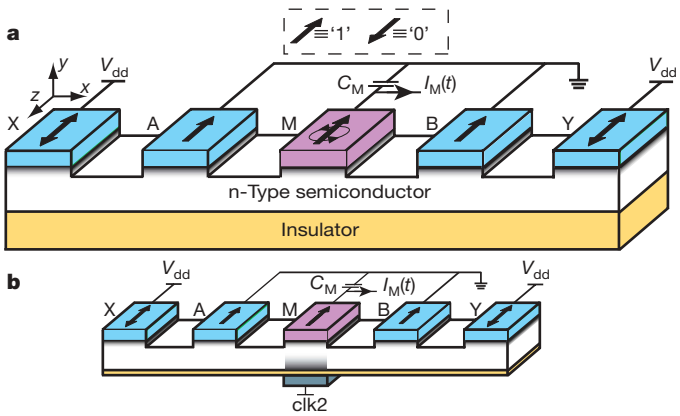


Figure 1 | Design of the reprogrammable magnetologic gate. The inputs are the magnetization directions of the contacts labelled by A, B, X and Y. The general boolean expression of the gate is: OR(XOR(A,X),XOR(B,Y)). One may use the magnetization alignments of A and B to set different logic operations. For example, the figure shows a universal NAND operation between X and Y (A = B = '1'). **a**, Steady-state currents, driven by V_{dd}, flow between A(B) and X(Y). The output is given by a transient current response, $I_M(t)$, caused by an in-plane single rotation of M. The amplitude of $I_M(t)$ is proportional to the spin accumulation in the semiconductor. In order to write the magnetization directions of the contacts in cascaded magnetologic gates, we first convert $I_M(t)$ to a binary voltage signal (see Fig. 3). This voltage signal will control the direction of a second, independent high-level current in wires on top of, and/or below, the contact(s) of other gate(s). **b**, A similar structure but with a pinned middle contact. The transient current, $I_M(t)$, is triggered by a voltage signal, clk2, applied to a (non-magnetic) back-contact beneath the mid-section. Semiconductor regions beneath the contacts are heavily doped.

[1]Department of Physics, University of California San Diego, La Jolla, California 92093-0319, USA.

gate: between A(B) and X(Y). In the parallel case, the excess injected spin population is easily drained from the channel, whereas in the antiparallel case opposite spins are more efficiently injected and extracted, leading to much larger spin accumulation[26]. The electron transport inside the semiconducting layer is described by the electrochemical potential, ζ_s^{sc}, defined as the sum of the electrical and chemical potentials, for each spin component $s = \pm$. Figure 2a shows typical steady state profiles of the electrochemical potentials for NAND operation with zero operands (A = B = '1', X = Y = '0'). The amount of spin accumulation is proportional to the potential-splitting. Similarly to the case of all-metallic non-local spin valves[13,14], these profiles are insensitive to the magnetization direction of M, being determined by the contacts A, B, X and Y, which inject and extract the steady state current. However, in our case a transient current is allowed to pass through the probing contact (M) during the magnetization rotation. The spin dependent conductances $G_+(t)$ and $G_-(t)$ of the Schottky barrier between M and the semiconductor follow the precession, and when the magnetization is reversed, the two spin conductances exchange roles. This gives rise to a transient current during which the potential level inside M is perturbed and the capacitor C_M (see Fig. 1) tries to restore new steady state conditions ($I_M = 0$). The magnetization dynamics of M then provides a means to electrically read out the result of the logic operation. During the transient, the peak current is proportional to the spin accumulation densities beneath the M contact.

The transient current $I_M(t)$ is shown in Fig. 2b in three different curves for four possible magnetization alignments of X and Y, corresponding to different logic inputs. (For '10' and '01' the results are the same, by the mirror symmetry about M.) The output signal corresponding to NAND(X = '1',Y = '1') = 0 should be well separated from the output signals corresponding to

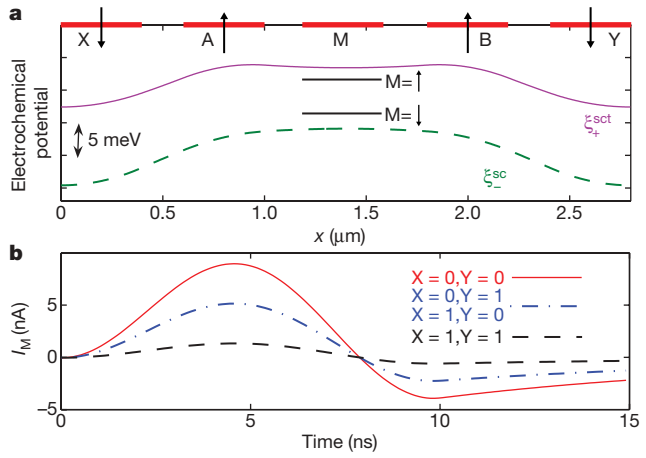


Figure 2 | Modelled electrical behaviour of a magnetologic gate set for NAND(X,Y). The diffusive transport in the planar semiconductor is modelled using a method in which the effects of traversing under the finite width of the metal contacts, and of the intrinsic capacitance across the Schottky barrier, are included in the regular spin diffusion formalism (see Supplementary Information). **a**, Typical steady state profiles of the spin dependent electrochemical potentials in the semiconductor channel for antiparallel configuration. Red sections denote the position of 0.4-μm-wide contacts, which are separated by 0.2-μm uncapped regions. In a steady state when $I_M = 0$, the potential level inside contact M for each magnetization alignment is represented by one of the two short black lines. **b**, Transient currents in the M terminal induced by a 10-ns rotation of M magnetization. The bias is $V_{dd} = 0.5$ V, the external capacitance is $C_M = 40$ fF, the conductive channel thickness is 100 nm and the contacts are 2 μm long. The semiconductor is 10^{16} cm^{-3} n-type GaAs at room temperature. The mobility and spin relaxation time are, respectively, 5,000 cm^2 V^{-1} s^{-1} and 80 ps (ref. 30). We use experimentally verified Fe/GaAs Schottky barrier parameters[5]: the heavy doping of the interfaces results in ~10-nm-thick barriers whose total conductance $(G_+ + G_-)$ is 1,000 Ω^{-1} cm^{-2}, intrinsic capacitance per unit area is ~10^{-6} F cm^{-2} and spin selectivity $|G_+ - G_-|/(G_+ + G_-) = 1/3$.

NAND(0,0) = NAND(0,1) = NAND(1,0) = 1. In order to maintain signal integrity, the noise level should be small so that the analysing circuit can distinguish between '0' and '1' logical outputs. In the bandwidth of interest (0.1 GHz, given by the inverse of the transient time), the mean square-root of the shot noise current and of the Johnson thermal noise current are, respectively, <0.5 nA and ~3.6 nA. The three currents corresponding to logical '1' are thus above the noise level whereas the logical '0' current is below it. In the Supplementary Information we discuss different ways to improve the signal integrity while reaching higher operation speed.

In order to build a competitive spin-based computer, we introduce a scheme to integrate the proposed MLGs into a large circuit. Advanced programmable CMOS logic circuits contain ~10^6 gates, which rival supercomputers for certain types of applications[27]. About half of the hardware in such complex circuits is needed to synchronize between dependent logic operations. The fundamental building block for synchronization is either the flip-flop or the latch, both of which store a bit of information. In a large circuit, each MLG is connected with a thyristor latch which converts a small transient current into a steady voltage signal. The internal structure of this latch is explained in the Supplementary Information. Here we explain how the latch manages the flow of information between sequential logic operations within three phases in time. In the first phase, each latch is prepared to receive a signal. In the second phase, the magnetization of the middle ferromagnetic contact in each MLG is rotated and the transient current signal (Fig. 2b) is captured by the associated thyristor latch. During the third phase, the output from each latch is a binary voltage signal that is high or low, depending on the amplitude of the transient current. In this phase, these voltages are used to control the preparation of the operands for the next logic cycle. The simultaneous operation of all MLGs is made possible by the use of three periodic and global clock signals, each representing a different phase. In addition to the thyristor latch, we use CMOS pass gates (transmission gates) to steer a high level current above the ferromagnetic contacts. Ordinarily, a pass gate uses a voltage signal to control the flow of a second current signal through the bulk of the pass gate. Using binary voltage levels results in high or low current flow, which are denoted, respectively, by 'on' or 'off' states. Figure 3 illustrates a representative example in which an outcome of a logic operation from a MLG propagates into inputs of two similar gates. We use a single, global static current to polarize operands of many MLGs along its path. That is, the high level current, I_h, in Fig. 3 also crosses above cascaded MLGs other than those that are shown in the figure.

This method of connecting the MLGs has several benefits over conventional electronics. When a single MLG output has to be connected to several MLG inputs, we simply add these MLGs between the p3 pass gates (no fan-out; see Fig. 3 for pass-gate nomenclature). On the other hand, CMOS circuits would require transistors of increasing size, which increase the circuit's area, power and delay. These costs can be quite substantial for common CMOS circuits, such as a 2^n-bit binary shifter where only n input signals must charge or discharge $n2^{n+1}$ transistors. As shown by Fig. 3, the high-level current I_h remains constant between the logic units and so eliminates dynamical transmission line effects associated with long wires. The main dynamical power dissipation is due to discharging the self-capacitance of three pairs of pass gates for each logic operation (the pass gates control the paths through which I_h propagates). The static power dissipation in our logic is caused by passing I_h through the leaky pass gates. It is assumed that the metallic wires between the semiconductor pass gates are short: they have less resistance and negligible time delay. The static power dissipation is kept relatively low by using short and wide channel pass gates, in which the resistance in the 'on' mode is $\leq 1\,\Omega$. These devices are built with a large leakage current in the 'off' state which, in contrast to CMOS, does not hinder the performance of the magnetoelectronics scheme. The Supplementary Information contains a detailed discussion of all

sources of power dissipation. Here we remark that the resulting power dissipation due to a million synchronized MLGs with I_h = 7 mA remains comparable to the power budget of a single, modern microprocessor (~50 W).

Future versions of the MLG may capitalize on both planar dimensions, allowing very complex logic operations every cycle. For example, one may lay down two additional ferromagnetic contacts on each side of the middle contact along the z axis of Fig. 1a. In such a clustered topology, eight operands are involved in a single reprogrammable logic operation. Additionally, the ability to reprogram the MLGs at every operational cycle increases their complexity without increasing the overall propagation delay or the number of gates needed to implement a demanding logic function. CMOS logic circuits cannot afford such flexibility without sacrificing either circuit area or speed. For example, field programmable gate arrays (FPGAs) use significantly more hardware to implement equivalent logic and practically can only be programmed at boot-up. The simulated 5.6 μm^2 planar MLG is comparable in size to the typical NAND gate in advanced CMOS-based microprocessors. The MLG may scale down considerably before its contacts reach the super-paramagnetic regime[28]. In contrast, planar CMOS transistors are now close to their intrinsic limits. Finally, the overhead in the proposed design, one thyristor latch per MLG, would not preclude its use in microprocessors in which only a few logic operations take place between latches. Because synchronization timing does not scale as fast as CMOS transistor speeds, the number of sequential logic operations continues to decline in microprocessors.

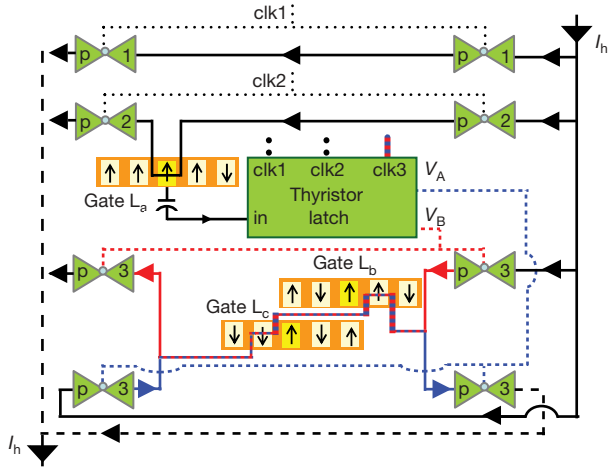


Figure 3 | Proposed logic cascading scheme. The output from a magnetologic gate, L_a, propagates into two similar gates, L_b and L_c. Lines with arrows denote current-carrying wires. Other lines hold binary voltage signals, which help to synchronize the flow of information and also steer the high level current I_h between the pass gates labelled by p1, p2 and p3. The synchronization process is managed in three phases by the thyristor latch (see text and Supplementary Information). In the first phase, the binary voltage signals are low except for clk1 at two inputs (dotted lines). This signal synchronizes two operations: one to prepare the thyristor latch to receive a signal and one to switch the pass gates p1 to their low resistance mode, thus directing the high current I_h to flow primarily in the upper horizontal black wire. During the second phase, clk2 is high and I_h flows primarily in the wire over the top of the middle contact of L_a. The locally induced magnetic field rotates the magnetization of this contact and thus triggers the transient current labelled by 'in'. The thyristor latch is synchronized by clk2 to capture this response. In the third phase, clk3 is high and one of the voltage signals V_A or V_B at the output of the thyristor latch is also high. Whether V_A or V_B is high depends on the amplitude of the transient current. This directs I_h into one of two opposite paths (respectively, blue or red wires), resulting in magnetic encoding of '0' or '1' in the relevant operands (A, B, X or Y) of L_b and L_c. Although not shown, during the three-phased action each magnetologic gate sends its output and then receives the four inputs for the next three-phase cycle.

Computing devices quickly evolved to use semiconductors because carrier densities could be widely controlled over orders of magnitude. This produced fast devices with low bit error rate (BER). Comparable magnetologic schemes, which involve switching the magnetization of 10^6 ferromagnetic contacts on each cycle, must achieve a competitive BER. Such a requirement is more stringent than for MRAM, in which only a few data bits are addressed in a given clock cycle. However, given the design of the magnetologic scheme (Fig. 3) and its departure from conventional MRAM design, most of the errant magnetization dynamics are eliminated (see Supplementary Information). Scaling down the size of the MLG may introduce a new BER source. The proximity of the operand contacts to the M writing lines may affect the logic evaluation during the rotation of M (clk2). The induced magnetic field may perturb the magnetization of A or B, changing the semiconductor's spin accumulation. As a result, the I_M signal may be suppressed or falsely amplified, leading to a latching error followed by a writing error in the cascaded gate. If the physical design exhibits this property, a possible solution could be the structure shown in Fig. 1b. In this configuration, the magnetization of M is pinned, and instead of using a magnetic field one has to apply the voltage clock signal, clk2, to the non-magnetic contact beneath the mid-section. The essential physics in a semiconductor sandwiched between a ferromagnet and a non-magnetic contact has recently been described for spin extraction[29]. Here, we highlight the features important for the present case.

The underlying contact in Fig. 1b controls the shape of the conduction band near the middle semiconductor/ferromagnet (S/F) interface. When clk2 is not active, electrons are free to move in the space between the middle S/F interface and the insulator. Free electrons, whose spins are aligned with M's majority spin population, tunnel more easily to and out of M. On the other hand, when clk2 is active, the biased contact depletes free electrons near the insulator interface (as in a regular field effect transistor). The electrons are localized in surface bands between the S/F Schottky barrier and the biased back-gate depleted region. Contrary to the case of free electrons, localized electrons whose spins are aligned with the minority spin population of M tunnel more easily into and out of the magnet. The switching of roles between the transport of free and of localized electrons yields the same behaviour as exchanging roles between the spin-up and spin-down conductances through the barrier. Thus, the transient current, I_M, is generated analogously as before.

In conclusion, we have proposed a unique computation scheme based on a system of ferromagnetic contacts on top of n-type semiconductor. We have presented the operation of a reprogrammable, universal spintronics gate and introduced a method to scale these magnetologic components to a very large size while respecting power constraints. These ingredients are sufficient for realization of any finite state machine (for example, a spin-computer). In addition, we have suggested an extension of the magnetologic gate, capitalizing on both spin accumulation and the electrical field effect in semiconductors.

Received 17 July 2006; accepted 10 April 2007.

1. Wolf, S. A. *et al.* Spintronics: A spin based electronics vision for the future. *Science* **294**, 1488–1495 (2001).
2. Ohno, Y. *et al.* Electrical spin injection in a ferromagnetic semiconductor heterostructure. *Nature* **402**, 790–792 (1999).
3. Dietl, T., Ohno, H., Matsukura, F., Cibert, J. & Ferrand, D. Zener model description of ferromagnetism in zinc-blende magnetic semiconductors. *Science* **287**, 1019–1022 (2000).
4. Zhu, H. J. *et al.* Room-temperature spin injection from Fe into GaAs. *Phys. Rev. Lett.* **87**, 016601 (2001).
5. Hanbicki, A. T. *et al.* Analysis of the transport process providing spin injection through a Fe/AlGaAs Schottky barrier. *Appl. Phys. Lett.* **82**, 4092–4094 (2003).
6. Jiang, X. *et al.* Highly spin-polarized room-temperature tunnel injector for semiconductor spintronics using MgO(100). *Phys. Rev. Lett.* **94**, 056601 (2005).
7. Crooker, S. A. *et al.* Imaging spin transport in lateral ferromagnet/semiconductor structures. *Science* **309**, 2191–2195 (2005).
8. Lou, X. *et al.* Electrical detection of spin transport in lateral ferromagnet–semiconductor devices. *Nature Phys.* **3**, 197–202 (2007).

9. Kato, Y. K., Myers, R. C., Gossard, A. C. & Awschalom, D. D. Observation of the spin Hall effect in semiconductors. *Science* **306,** 1910–1913 (2004).
10. Wunderlich, J., Kaestner, B., Sinova, J. & Jungwirth, T. Experimental observation of the spin-Hall effect in a two-dimensional spin-orbit coupled semiconductor system. *Phys. Rev. Lett.* **94,** 047204 (2005).
11. Baibich, M. N. *et al.* Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices. *Phys. Rev. Lett.* **61,** 2472–2475 (1988).
12. Binasch, G., Grünberg, P., Saurenbach, F. & Zinn, W. Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange. *Phys. Rev. B* **39,** R4828–R4830 (1989).
13. Johnson, M. Bipolar spin switch. *Science* **260,** 320–323 (1993).
14. Jedema, F. J., Filip, A. T. & van Wees, B. J. Electrical spin injection and accumulation at room temperature in an all-metal mesoscopic spin valve. *Nature* **410,** 345–348 (2001).
15. Stephens, J. *et al.* Spin accumulation in forward-bias MnAs/GaAs Schottky diodes. *Phys. Rev. Lett.* **93,** 097602 (2004).
16. Johnson, M. in *Magnetoelectronics* (ed. Johnson, M.) 273–330 (Elsevier, San Diego, 2004).
17. Cowburn, R. P. & Welland, M. E. Room temperature magnetic quantum cellular automata. *Science* **287,** 1466–1468 (2000).
18. Hanbicki, A. T. *et al.* Nonvolatile reprogrammable logic elements using hybrid resonant tunneling diode–giant magnetoresistance circuits. *Appl. Phys. Lett.* **79,** 1190–1192 (2001).
19. Richter, R., Bär, L., Wecker, J. & Reiss, G. Non-volatile field programmable spin-logic for reconfigurable computing. *Appl. Phys. Lett.* **80,** 1291–1293 (2002).
20. Ney, A., Pampuch, C., Koch, R. & Ploog, K. H. Programmable computing with a single magnetoresistive element. *Nature* **425,** 485–487 (2003).
21. Imre, A. *et al.* Majority logic gate for magnetic quantum-dot cellular automata. *Science* **311,** 205–208 (2006).
22. Datta, S. & Das, B. Electronic analogue of the electro-optic modulator. *Appl. Phys. Lett.* **56,** 665–667 (1990).
23. Kikkawa, J. M. & Awschalom, D. D. Resonant spin amplification in n-type GaAs. *Phys. Rev. Lett.* **80,** 4313–4316 (1998).
24. Tehrani, S. *et al.* Recent developments in magnetic tunnel junction MRAM. *IEEE Trans. Magn.* **36,** 2752–2757 (2000).
25. Gerrits, Th, van den Berg, H. A. M., Hohlfeld, J., Bär, L. & Rasing, Th Ultrafast precessional magnetization reversal by picosecond magnetic field pulse shaping. *Nature* **418,** 509–512 (2002).
26. Dery, H., Cywiński, Ł. & Sham, L. J. Spin transference and magnetoresistance amplification in a transistor. *Phys. Rev. B* **73,** 161307(R) (2006).
27. Hauck, S. The roles of FPGAs in reprogrammable systems. *Proc. IEEE* **86,** 615–639 (1998).
28. Skumryev, V. *et al.* Beating the superparamagnetic limit with exchange. *Nature* **423,** 850–853 (2003).
29. Dery, H. & Sham, L. J. Spin extraction theory and its relevance to spintronics. *Phys. Rev. Lett.* **98,** 046602 (2007).
30. Pikus, G. E. & Titkov, A. N. in *Optical Orientation* Vol. 8 (eds Meier, F. & Zakharchenya, B. P.) 73–131 (North-Holland, New York, 1984).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Dalal for suggestions regarding the logic cascading scheme. This work was supported by the National Science Foundation.

Author Contributions H.D. and L.J.S. developed the proposed idea of spin computation. P.D. designed the cascading scheme. Ł.C. formulated the time-dependent spin-diffusion transport.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.D. (hdery@ucsd.edu).

LETTERS

Observational evidence for an ocean heat pump induced by tropical cyclones

Ryan L. Sriver[1] & Matthew Huber[1,2]

Ocean mixing affects global climate and the marine biosphere because it is linked to the ocean's ability to store and transport heat[1] and nutrients[2]. Observations have constrained the magnitude of upper ocean mixing associated with certain processes[3,4], but mixing rates measured directly[3,5] are significantly lower than those inferred from budget analyses[6], suggesting that other processes may play an important role. The winds associated with tropical cyclones are known to lead to localized mixing of the upper ocean[7–9], but the hypothesis that tropical cyclones are important mixing agents at the global scale[10] has not been tested. Here we calculate the effect of tropical cyclones on surface ocean temperatures by comparing surface temperatures before and after storm passage, and use these results to calculate the vertical mixing induced by tropical cyclone activity. Our results indicate that tropical cyclones are responsible for significant cooling and vertical mixing of the surface ocean in tropical regions. Assuming that all the heat that is mixed downwards is balanced by heat transport towards the poles, we calculate that approximately 15 per cent of peak ocean heat transport may be associated with the vertical mixing induced by tropical cyclones. Furthermore, our analyses show that the magnitude of this mixing is strongly related to sea surface temperature, indicating that future changes in tropical sea surface temperatures may have significant effects on ocean circulation and ocean heat transport that are not currently accounted for in climate models.

The ocean circulation is complex, time-dependent and three-dimensional, but it is generally accepted that the meridional overturning circulation (MOC) is a key climate feature because it is linked with ocean heat transport (OHT), and the MOC is itself maintained by mechanical diapycnal mixing[1,11]. Below the near-surface mixed layer and above the main thermocline base, this mixing is usually driven by shear instability[4,12]. Internal waves, driven either by winds or tides, are the main sources of shear[1]. As these waves break, they generate most of the upper ocean's diapycnal mixing[4]. Winds drive at least half of the total mixing[1], with some studies giving a leading role to wind-driven mixing in the Southern Ocean[13] and some preferring a tropical location[14]. Both sources of mixing may be important. Budget analyses require 16 and 20 Sv ($1\,\mathrm{Sv} = 10^6\,\mathrm{m^3\,s^{-1}}$) of diapycnal cross-thermocline upwelling to occur within the tropical-to-midlatitude Pacific[15] and Atlantic[16], respectively, in addition to substantial Southern Ocean mixing. Furthermore, tropical upper ocean mixing is considered especially important to climate because simulations and theory agree that this region of strong stratification is the most efficient place for mixing to drive strong heat transport[14,17]—climate is especially sensitive to mixing variations in this region[18].

The magnitude and causes of upper ocean mixing are not well established, and the spread of values between many low measured values and those required by macroscopic balance requirements remains large[1]. Theory and microstructure-based measurements agree in that they both suggest an extremely weak low latitude mixing coefficient ($0.005–0.05\,\mathrm{cm^2\,s^{-1}}$), but the values are generally higher and range by a factor of 30 in subtropics and midlatitudes[3]. Upper ocean tracer release experiments generally agree with these estimates, showing a vertical mixing coefficient of $\sim$0.1–$0.2\,\mathrm{cm^2\,s^{-1}}$ (ref. 5) in most locations but also with substantial variability[19]. Other approaches using large-scale energy budget requirements[6] indicate time-mean and basin-wide diffusivities at the higher end ($0.1–1.5\,\mathrm{cm^2\,s^{-1}}$). Upper ocean background diffusivities of $\sim0.1\,\mathrm{cm^2\,s^{-1}}$ are required in global ocean models. Reconciling these different approaches and results is difficult, but differences in approach between the methods provide insights. Upper ocean studies that have only been carried out for brief intervals (days to weeks) or that have sampled the oceans sparsely (the majority of microstructure and dye tracer studies) tend to show smaller values. Budget studies that integrate over years and over entire water masses tend to show higher values, suggesting that differences between these approaches might be due to spatially or temporally localized phenomena. Consequently, transient and localized strong sources of internal waves may provide a resolution to remaining discrepancies. Recent results indicate that transient, localized mixing—even in the tropics—may contribute significantly to the MOC, as long as the mixing happens near ocean boundaries[20].

Tropical cyclones (hereafter 'cyclones') are transient events, and their high wind speeds generate strong, near-inertial internal waves, making them efficient upper ocean mixers[7]. This diapycnal mixing is evidenced by the prominent cold wakes seen along, and to the right of, storm tracks, with sea surface temperature (SST) depressed by up to 8 °C (Supplementary Fig. S1). The wakes are primarily (>75%) due to upward mixing of cold water caused by the breaking of inertial waves that entrain water from the mixed-layer base[7–9]. This mixing is known to transport heat downwards[10,21] and nutrients upwards[2]. Cold wakes last days to weeks, and surface conditions are restored to climatologically normal conditions through surface fluxes[10]. The downward-mixed heat anomaly persists, and the full vertical column experiences a net warming[10,21].

This downward pumping of warm water by cyclones represents a heat convergence—that is, a net increase in ocean heat content (OHC). In steady state, such heat convergence should be balanced by a lateral transport of ocean heat out of the storm-affected region. Emanuel[10] used observations of cyclone cold wakes, together with model simulations using a simple coupled, mixed layer ocean model, to estimate cyclone-induced OHT for 1996. His estimate is 1.4 petawatts ($1\,\mathrm{PW} = 10^{15}\,\mathrm{W}$) with a total uncertainty of ±0.7 PW, which is sufficient to account for the peak observed OHT in the subtropics. Emanuel[22] also hypothesized that the integrated intensity of cyclone events might be linked to the mean climate state by the potential relationship between cyclone activity and SST[23,24]. Thus, feedbacks could exist between cyclones, ocean heat convergence and transport, profoundly altering the behaviour of the climate system with higher-than-modern temperatures.

[1]Department of Earth and Atmospheric Sciences, [2]Purdue Climate Change Research Center, Purdue University, West Lafayette, Indiana 47907, USA.

We test the key elements of this hypothesis by comparing surface temperatures before and after storm passage, using reanalysis data. The methodology builds on the techniques described in ref. 24 (see Methods). We find that cyclones regularly cool the tropical oceanic mixed layer (Fig. 1a), generate strong vertical mixing (Fig. 1b), and pump downward the heat lost from the mixed layer (Fig. 1c). These processes are a strong function of tropical SST (Fig. 2). The tropical SSTs are depressed by up to several degrees climatologically where tropical cyclone activity is greatest (Fig. 1a). This temperature anomaly pattern is found in multiple surface temperature data sets (Supplementary Fig. 2). Phrased as a W m^{-2} forcing, the upper ocean is cooled by ~1–3 W m^{-2} (Fig. 1a) annually, assuming that the SST anomaly penetrates to 50 m depth. Interestingly, the locations for this heat input are well correlated with observed OHC anomalies, suggesting a positive correlation between upper ocean heat content changes, cyclone activity, and cyclone-induced heat pumping (Fig. 1c), as required by Emanuel's hypothesis.

Using observed, spatially resolved, cyclone-season averaged, ocean vertical temperature profiles[25], we estimate the depth of cyclone-induced mixing (Supplementary Fig. S3). This mixing depth length scale is used to express the cyclone-induced cooling as an effective vertical diffusivity (Fig. 1b) (see Methods). We use diapycnal, vertical and diathermal diffusivity interchangeably because they are generally equivalent in the regions of interest[1,6]. This map of annualized vertical diffusivity clearly shows maxima in regions with substantial cyclone activity, and much of the mixing occurs near western boundaries. The values vary from close to the minimum values observed from direct measurements (~0.05 cm^2 s^{-1}) and up to much higher values (>1 cm^2 s^{-1}) in the western Pacific. Nearly all tracer and microstructure studies have avoided measuring cyclone-influenced conditions. The one published measurement for cyclone-influenced conditions is 10 times the undisturbed conditions measured at that location[19], and agrees with our calculation. The lack of cyclone-induced mixing near the Equator is consistent with weak mixing noted there[3].

Comparison with integral budget-based values[6] reveals remarkably close agreement, both in terms of inferred global, tropical mean diffusivity and their sharp increases over the warmest ocean water masses (Fig. 2a). The close agreement of these data sets can be parsimoniously interpreted to indicate that much or possibly all of the upper ocean mixing observed in these analyses[6] is attributable to cyclone activity. These results support the conjecture that tropical uppermost ocean mixing is quite weak except during the occasional cyclone mixing event[12]. This could have serious implications for climate models, which represent this mixing as a constant 'background' diffusivity, and also for other analyses dependent on vertical mixing, such as models for primary production.

We calculate cyclone-induced OHC changes by assuming that surface temperature anomalies indicate uniform cooling to 50 m depth, and that this heat lost from the mixed layer is pumped downward. We represent the annualized increase in OHC as the steady-state cyclone-induced global OHT, assuming that all heat converged must, in steady state, become a poleward transport. This value for OHT estimated from reanalysis is 0.26 PW, with peak values greater than 0.50 PW (Supplementary Fig. S4). Our results indicate that ~15% of peak global OHT may currently be directly related to cyclone-derived mixing. This value is comparable to the total transport in the Atlantic Ocean past 50° N (ref. 26). These values are

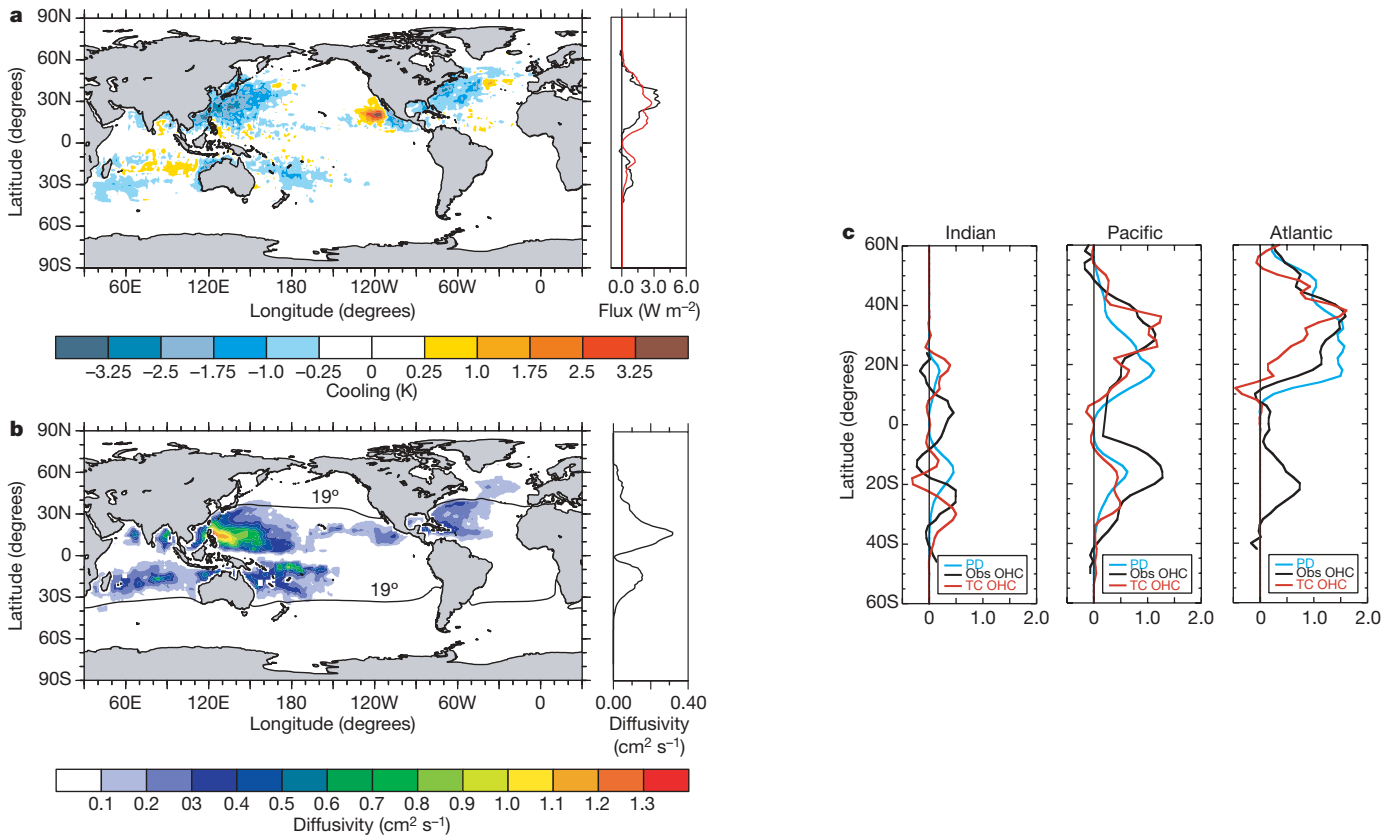


Figure 1 | Maps of tropical cyclone effects on the upper ocean. **a**, Average cooling for the annually accumulated, cyclone-induced temperature anomalies derived from ERA-40 2 metre air temperature (2MT; 1982–2001). Inset, the zonally averaged fluxes needed to restore the heat anomalies over one year, assuming the anomalies are 50 m deep (black line, 2MT; red line, SST; see Methods and Supplementary Information). **b**, Annualized average of vertical diffusivity attributable to cyclone mixing (see Methods). Contour represents the 19 °C isotherm, the margin of the 'warm water sphere'. Inset, the zonally averaged diffusivity. **c**, Annually averaged zonal-mean cyclone power dissipation (PD; blue curves), observed OHC anomalies from IGOSS (Obs OHC; black curves), and tropical-cyclone-induced OHC anomalies (TC OHC; red curves) for the period 1994 to 2001. Observed OHC curves are divided by 4×10^8 J m^{-2}. Dissipation and cyclone OHC are normalized by the respective observed OHC for each basin.

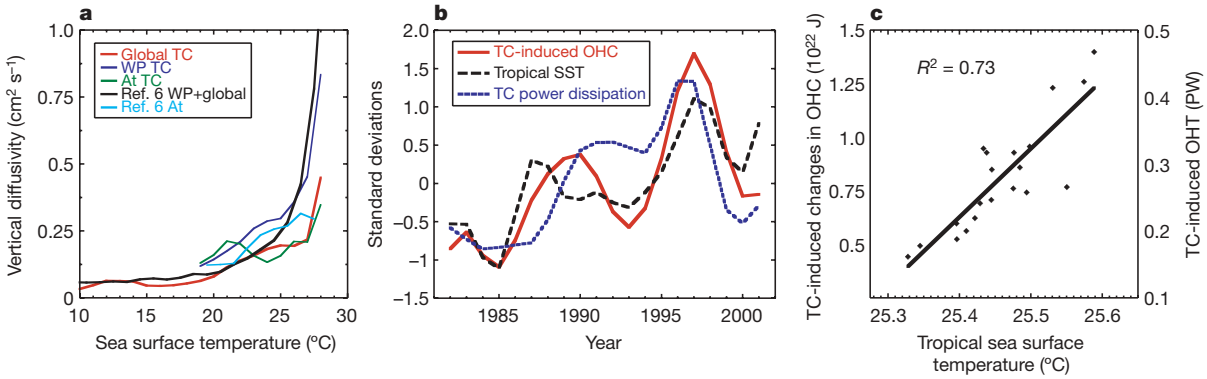


Figure 2 | Potential cyclone-induced climate interactions with vertical ocean mixing. **a**, Effective vertical diffusivity versus SST for global tropical-cyclone mixing (Global TC; red line), western north Pacific tropical-cyclone mixing (WP TC; dark blue line), and north Atlantic tropical-cyclone mixing (At TC; green line)—compared to global (WP+global; black line) and Atlantic (At; light blue line) diffusivity observations from ref. 6. **b**, Annual tropical-cyclone-induced OHC changes derived from ERA-40 2 metre air temperature (2MT) (solid, red curve), mean annual tropical SST from ERA-40 (dashed, black curve), and globally integrated tropical-cyclone power dissipation (PD) from ERA-40 near surface winds (dotted, blue curve). All data have been 5-yr low-pass-filtered, and are normalized by their respective standard deviations. **c**, Scatter plot of tropical-cyclone-induced OHC anomalies (left axis) and the equivalent OHT (right axis) derived from ERA-40 2MT and mean annual tropical SST from ERA-40. Both quantities are filtered as in Fig. 2b.

significantly smaller than early estimates[10], but are consistent with recent modelling work[27].

More important than this quantity's current mean value is its potential climate sensitivity, because previous studies observed correlations between increasing cyclone activity and rising SST[23,24]. Figure 2b shows the time series from 1982 to 2001 of globally integrated and annually averaged quantities as follows: cyclone-induced increases in OHC, tropical SST, and cyclone power dissipation (an integrated measure of cyclone wind intensity, hereafter referred to as dissipation). Tropical SST and globally integrated annual OHC/OHT correlate well, with $r^2 = 0.73$ (Fig. 2c). This agreement is consistent with theory[22], and suggests that cyclones may play a key role in climate dynamics. Close agreement also exists between integrated dissipation and cyclone-induced OHT ($r^2 = 0.57$, Supplementary Fig. S5), which is consistent with the hypothesis that increases in the intensity, frequency or duration of intense storms should lead to more extensive, vigorous mixing and more pronounced cold wakes. A 0.25 °C increase in mean annual tropical SST may lead to a 60% increase in global dissipation[24]—we show here that this SST increase may lead to a tripling of cyclone-induced global OHT (Fig. 2c).

As a consistency check on these dissipation calculations, we compute the power available for turbulent mixing and compare this with independent estimates. The average total reanalysis-derived dissipation observed through this interval (3.3×10^{19} J) corresponds to 1×10^{12} W expended on turbulent mixing. Using different methods, the global contribution of cyclones to the ocean's near-inertial spectral power range is calculated to be $\sim 7 \times 10^{11}$ W (ref. 28), consistent with our estimate. This is fully half of the total mixing required[1] to balance 30 Sv of deep-water formation, so it appears that mechanical stirring by cyclones may be responsible for about half of what is commonly called the thermohaline circulation. Furthermore, the results are consistent with—and provide a physical explanation for—water-mass budgets requiring tropical cross-thermocline mixing[15,16]. The importance of the cyclone-induced contribution has hitherto gone unnoticed in some previous studies[11], probably because wind fluctuations with timescales less than 5 days were not considered—hence the cyclone contribution was filtered out. A preliminary analysis indicates that even wind data sets with daily resolution that are used to drive ocean models underrepresent the integrated wind intensity by a factor of 10, so no current simulations are likely to accurately capture this highly nonlinear mixing process. This may partially explain a frequently noted tendency of ocean models to produce overly shallow mixed layer depths.

Our analysis suggests that changes in global cyclone frequency, duration and/or intensity are closely related to the amount of heat pumped into—and available to be subsequently transported by—the oceans. This relationship may have implications for changes in heat transport associated with past and future climate change. Extrapolation of our results suggests that future increases in tropical temperatures may result in increased dissipation, mixing, heat storage, and eventually heat transport. Moreover, this positive response in transport might feed back on climate by redistributing heat poleward, diminishing the Equator-to-pole temperature gradient, and raising global mean temperature[29]. We have provided some evidence that cyclone-induced mixing is a fundamental physical mechanism that may act to stabilize tropical temperatures, mix the upper ocean, and cause polar amplification of climate change. It is not included in the current conceptual or numerical models of the climate system. Better representation of cyclone winds and the associated mixing in climate models may help to explain the still-vexing questions posed by past climates[30].

METHODS

Data. We primarily show results for two-metre air temperature (2MT) from the European Centre for Medium-Range Weather Forecasts Reanalysis Project (ERA-40) after 1981. Additional results from other data sources are discussed in the Supplementary Information. We explicitly assume that the cyclone-induced temperature depression is due entirely to vertical mixing (see Supplementary Information). We use cyclone tracks from the 'best track' data sets (see ref. 24 for details) and SSTs at 6 h intervals over a $6^\circ \times 6^\circ$ footprint centred on the storm's eye.

Calculation of cyclone-induced temperature anomalies. To calculate anomalies, final temperatures are taken at each location 3 days after storm passage for 2MT fields and 7 days after for SST and skin temperature (SKT) fields; final values are subtracted from the initial temperature conditions 3 days before the storm for all fields. Separate anomaly timescales are used for measuring 2MT, SST and SKT because the characteristic variability of each variable within the reanalysis data sets varies. SST data after 1981 within reanalysis are based on weekly averages, so that a longer timescale is needed in order to capture the cyclone-induced anomaly compared with 2MT. On the other hand, 2MT is more variable owing to atmospheric processes, and we use a shorter timescale for reanalysis fields for 2MT anomalies (6 days) with respect to SST anomalies (10 days).

Calculation of diffusivity attributable to cyclone mixing. We calculate the effective vertical diffusivity, k_v, as $k_v = L^2/\tau$, where L is a depth scale over which mixing occurs and τ is a characteristic timescale over which the mixed layer deepening and entrainment occur. We use 'vertical' here, but diapycnal or diathermal would be equivalent statements, as described earlier. To calculate the effective vertical diffusivity attributable to cyclone mixing (Fig. 1b), we begin by assuming that all mixing in a given year is achieved during the single largest

cooling event, that is, the largest cyclone-induced temperature anomaly at each location is indicative of all the mixing at that location during a given year. This assumption appears reasonable, given the small diffusivities measured in tropical regions[3] in undisturbed conditions, and is consistent with the observation that all effective mixing in these locations is attributable to a few strong mixing events per year[12].

We estimate L using storm-season-averaged vertical temperature profiles ($\partial T/\partial z$) from ref. 25 in tandem with the SST anomalies (ΔT) from ERA-40. This technique assumes that the anomaly measured at the surface reflects well-mixed conditions down to some depth. We define L at each location as the level from which upwelling needs to occur in order to achieve the observed surface temperature anomaly, $L = \Delta T(\partial z/\partial T)$. We estimate that τ is $\sim$24 h, based on analysing individual storm track anomalies using satellite-based estimates with high temporal resolution (available at www.ssmi.com).

Calculation of cyclone-induced changes in OHC. The vertically integrated heat anomaly, Q, is calculated, adapting the formalism of Emanuel[10], as:

$$Q = \iiint F\rho C\Delta T \mathrm{d}h\mathrm{d}W\mathrm{d}S$$

where F is the fraction of heat transported downward from the oceanic mixed layer, ρ and C are respectively the density and heat capacity of sea water, ΔT and $\mathrm{d}h$ are respectively the magnitude and the depth of the temperature anomaly, and $\mathrm{d}W$ and $\mathrm{d}S$ are respectively the cross-track length and the along-track length of the storm wake.

For all calculations, ρ and C are held constant and equal to 1,020 kg m^{-3} and 3,900 J kg^{-1} °C^{-1}, respectively. To simplify the depth of the vertical mixing ($\mathrm{d}h$), we assume the depths of all heat anomalies are constant and equal to 50 m. This value for $\mathrm{d}h$ is likely to be an underestimate for strong storms, for which vigorous vertical motions have been observed and modelled down to depths of 200 m. We assume $F = 1$, and thus, all heat lost from the oceanic mixed layer is transported downward and ultimately poleward.

Received 5 January; accepted 30 March 2007.

1. Wunsch, C. & Ferrari, R. Vertical mixing, energy, and the general circulation of the oceans. *Annu. Rev. Fluid Mech.* **36,** 281–314 (2004).
2. Lin, W. *et al.* New evidence for enhanced ocean primary production triggered by tropical cyclone. *Geophys. Res. Lett.* **30,** doi:10.1029/2003GL017141 (2003).
3. Gregg, M. C., Sanford, T. B. & Winkel, D. P. Reduced mixing from the breaking of internal waves in equatorial waters. *Nature* **422,** 513–515 (2003).
4. Alford, M. H. Redistribution of energy available for ocean mixing by long-range propagation of internal waves. *Nature* **423,** 159–163 (2003).
5. Ledwell, J. R., Watson, A. J. & Law, C. S. Evidence for slow mixing across the pycnocline from an open-ocean tracer-release experiment. *Nature* **364,** 701–703 (1993).
6. Schneider, E. K. & Bhatt, U. S. A dissipation integral with application to ocean diffusivities and structure. *J. Phys. Oceanogr.* **30,** 1158–1171 (2000).
7. Price, J. F. Upper ocean response to a hurricane. *J. Phys. Oceanogr.* **11,** 153–175 (1981).
8. Jacob, S. D., Shay, L. K., Mariano, A. J. & Black, P. G. The 3D oceanic mixed layer response to Hurricane Gilbert. *J. Phys. Oceanogr.* **30,** 1407–1429 (2000).
9. D'Asaro, E. A. The ocean boundary below Hurricane Dennis. *J. Phys. Oceanogr.* **33,** 561–579 (2003).
10. Emanuel, K. A. The contribution of tropical cyclones to the oceans' meridional heat transport. *J. Geophys. Res.* **106,** 14771–14782 (2001).
11. Dalan, F., Stone, P. H., Kamenkovich, I. V. & Scott, J. R. Sensitivity of the oceans' climate to diapycnal diffusivity in an EMIC. Part I: Equilibrium state. *J. Clim.* **18,** 2460–2481 (2005).
12. Raymond, D. J. *et al.* EPIC2001 and the coupled ocean-atmosphere system. *Bull. Am. Meteorol. Soc.* **85,** 1341–1354 (2004).
13. Naveira Garabato, A. C., Polzin, K. L., King, B. A., Heywood, K. J. & Visbeck, M. Widespread intense turbulent mixing in the Southern Ocean. *Science* **303,** 210–213 (2004).
14. Scott, J. R. & Marotzke, J. The location of diapycnal mixing and the meridional overturning circulation. *J. Phys. Oceanogr.* **32,** 3578–3595 (2002).
15. Nof, D. & Van Gorder, S. Upwelling into the thermocline of the Pacific ocean. *Deep-sea Res. I* **47,** 2317–2340 (2000).
16. Nof, D. & Van Gorder, S. A different perspective on the export of water from the south Atlantic. *J. Phys. Oceanogr.* **29,** 2285–2302 (1999).
17. McWilliams, J. C., Danabasoglu, G. & Gent, P. R. Tracer budgets in the warm water sphere. *Tellus A* **48,** 179–192 (1996).
18. Bugnion, V., Hill, C. & Stone, P. H. An adjoint analysis of the meridional overturning circulation in an ocean model. *J. Clim.* **19,** 3732–3750 (2006).
19. Oakey, N. S. & Greenan, B. J. W. Mixing in a coastal environment: 2. A view from microstructure measurements. *J. Geophys. Res.* **109,** C10014, doi:10.1029/2003JC002193 (2004).
20. Boos, W. R., Scott, J. R. & Emanuel, K. A. Transient diapycnal mixing and the meridional overturning circulation. *J. Phys. Oceanogr.* **34,** 334–341 (2004).
21. Zedler, S. E. *et al.* Analyses and simulations of the upper ocean's response to Hurricane Felix at the Bermuda testbed mooring site: 13–23 August 1995. *J. Geophys. Res.* **107,** doi:10.1029/2001JC000969 (2002).
22. Emanuel, K. A. A simple model of multiple climate regimes. *J. Geophys. Res.* **107,** doi:10.1029/2001JD001002 (2002).
23. Emanuel, K. A. Increasing destructiveness of tropical cyclones over the past 30 years. *Nature* **436,** 686–688 (2005).
24. Sriver, R. L. & Huber, M. Low frequency variability in globally integrated tropical cyclone power dissipation. *Geophys. Res. Lett.* **33,** L11705, doi:10.1029/2006GL026167 (2006).
25. Levitus, S., Antonov, J. I., Boyer, T. P. & Stephens, C. Warming of the world ocean. *Science* **287,** 2225–2229 (2000).
26. Ganachaud, A. & Wunsch, C. Large-scale ocean heat and freshwater transports during the world ocean circulation experiment. *J. Clim.* **16,** 696–705 (2003).
27. Korty, R. L., Emanuel, K. A. & Scott, J. R. Tropical cyclone-induced upper ocean mixing and climate: application to equable climates. *J. Clim.* (submitted).
28. Shay, L. K. & Jacob, S. D. Relationship between oceanic energy fluxes and surface winds during tropical cyclone passage. In *Atmosphere-Ocean Interactions* Vol. 2 (WIT Press, Southampton, UK, in the press).
29. Herweijer, C., Seager, R., Winton, M. & Clement, A. Why ocean heat transport warms the global mean climate. *Tellus A* **57,** 662–675 (2005).
30. Sluijs, A. *et al.* Subtropical Arctic ocean temperatures during the Palaeocene/Eocene thermal maximum. *Nature* **441,** 610–613 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Schneider and K. Emanuel for diffusivity values (used in Fig. 2a) and hurricane track data, respectively. ERA-40 data were provided by the Data Support Section of the Scientific Computing Division at the National Center for Atmospheric Research (NCAR). NCAR is supported by the NSF. NCEP reanalysis data were provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA, from their website at http://www.cdc.noaa.gov. TMI data are produced by Remote Sensing Systems and sponsored by the NASA Earth Science REASoN DISCOVER Project. Data are available at www.remss.com. NODC_WOA98 data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at http://www.cdc.noaa.gov. M.H.'s research is supported by the NSF, the Purdue Research Foundation, the Purdue Cyber Center, and Information Technology at Purdue (ITaP).

Author Contributions R.L.S. and M.H. contributed equally to the writing, data analysis and ideas in this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.H. (huberm@purdue.edu).

LETTERS

Life-history trade-offs favour the evolution of animal personalities

Max Wolf[1], G. Sander van Doorn[1]†, Olof Leimar[2] & Franz J. Weissing[1]

In recent years evidence has been accumulating that personalities are not only found in humans[1] but also in a wide range of other animal species[2–8]. Individuals differ consistently in their behavioural tendencies and the behaviour in one context is correlated with the behaviour in multiple other contexts. From an adaptive perspective, the evolution of animal personalities is still a mystery, because a more flexible structure of behaviour should provide a selective advantage[9–11]. Accordingly, many researchers view personalities as resulting from constraints imposed by the architecture of behaviour[7] (but see ref. 12). In contrast, we show here that animal personalities can be given an adaptive explanation. Our argument is based on the insight that the trade-off between current and future reproduction[13] often results in polymorphic populations[14] in which some individuals put more emphasis on future fitness returns than others. Life-history theory predicts that such differences in fitness expectations should result in systematic differences in risk-taking behaviour[15]. Individuals with high future expectations (who have much to lose) should be more risk-averse than individuals with low expectations. This applies to all kinds of risky situations, so individuals should consistently differ in their behaviour. By means of an evolutionary model we demonstrate that this basic principle results in the evolution of animal personalities. It simultaneously explains the coexistence of behavioural types, the consistency of behaviour through time and the structure of behavioural correlations across contexts. Moreover, it explains the common finding that explorative behaviour and risk-related traits like boldness and aggressiveness are common characteristics of animal personalities[2–8].

The phenomenon of animal personalities is one of the most intriguing challenges to the adaptationist programme in behavioural research. Empirical findings in more than 60 species, ranging from primates to ants, suggest that animal behaviour is much less flexible than previously thought[2–8]. Individuals consistently differ in whole suites of correlated behaviours and these differences are often heritable[16–19]. At present, the existence of such personalities (also termed behavioural syndromes[20], coping styles[5] or temperaments[21]) is puzzling in several respects. First, why do different personality types stably coexist? Second, why is behaviour not more flexible but correlated across contexts and through time? And third, why are the same types of traits correlated in very different taxa[5–7]? Here we develop an evolutionary model that provides answers to all of these questions.

We start with the observation that some of the most prominent personality traits described in the literature can be categorized in terms of risk-taking behaviour. A good example is the correlation between aggressiveness towards conspecifics and boldness towards predators: individuals that risk more in intraspecific fights also risk more when confronted with a predator. This aggression–boldness syndrome has been described for many species[7], including fish[22,23], birds[8] and rodents[5]. From life-history theory it is known that individuals should adjust their risk-taking behaviour to their residual reproductive value[13,15], that is, their expected future fitness. Individuals with relatively high expectations should be relatively risk-averse, because they have to survive to realize those expectations. By the same reasoning, individuals with relatively low expectations should be relatively risk-prone because they have little to lose. Consequently, whenever individuals differ in their fitness expectations, we should expect stable individual differences and correlated behavioural traits: some individuals are consistently risk-prone whereas others are consistently risk-averse.

By means of a simple model we now show that these intuitive arguments do indeed provide an evolutionary explanation for animal personalities. We proceed in three steps. First, we show that the trade-off between current and future reproduction can easily give rise to polymorphic populations in which some individuals put more emphasis on future reproduction than others. Second, we demonstrate that this variation in life-history strategies selects for systematic differences in risk-aversion. Third, we show that these differences in risk-taking behaviour extend to various risky situations and are stable over time, thereby giving rise to animal personalities.

Consider the following stylized life history (Fig. 1a). Individuals live for two years and reproduce at the end of each year. The foraging habitat is heterogeneous with both high- and low-quality resources. Individuals face a trade-off between reproduction in year 1 and reproduction in year 2 that is mediated by exploration behaviour. We characterize the exploration behaviour by the strategic variable x, which ranges from superficial ($x=0$) to thorough ($x=1$). Individuals that explore their environment thoroughly have a high probability of obtaining a high-quality resource in year 2. For simplicity, we let this probability correspond to x. Yet, the probability of reproducing in year 1, $g(x)$, decreases with the intensity of exploration. Here we take $g(x)=(1-x)^{\beta}$, where $\beta>1$.

The payoff from feeding on high- or low-quality resources declines with the density of individuals (N_{high} or N_{low}, respectively) competing for such resources. It is given by:

$$F_i = \frac{f_i}{1+\alpha N_i} \qquad (1)$$

for i = high or low, where $\alpha>0$ represents the strength of competition and f_{high} and f_{low} (where $f_{\text{high}}>f_{\text{low}}$) denote the intrinsic benefits of obtaining a high- and a low-quality resource, respectively. At the end of each year, individuals produce a number of offspring that is proportional to the payoff they obtained in that year. To summarize, an individual with exploration intensity x produces $g(x)F_{\text{low}}$ offspring at the end of its first year; at the end of its second year it produces F_{high} offspring with probability x and F_{low} offspring with probability $1-x$.

[1]Theoretical Biology Group, Centre for Ecological and Evolutionary Studies, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands. [2]Department of Zoology, Stockholm University, SE-106 91 Stockholm, Sweden. †Present address: Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA.

For this basic life cycle, natural selection gives rise to the stable coexistence of two extreme exploration strategies (Fig. 1b): some individuals explore the environment thoroughly, thereby investing in future reproductive success, whereas others explore superficially, putting more emphasis on current reproduction. This dimorphism is a stable evolutionary outcome whenever $2f_{low} > f_{high} > 1$ (see Supplementary Information).

We now extend the above life history by assuming that each individual is confronted with a number of risky contexts[20] throughout its lifetime (see Methods). Between year 1 and year 2, each individual may face one or more foraging decisions under predation risk (anti-predator games) and one or more aggressive encounters with randomly chosen conspecifics. In each of the anti-predator games an individual can either behave in a bold or a shy manner. Only bold individuals receive a payoff (in terms of higher fecundity) but they also incur some risk of dying. Aggressive encounters are modelled as hawk–dove games[24], in which hawks receive a higher payoff than doves but do not always survive hawk–hawk interactions.

Let us first consider the two most basic cases, where all individuals either play one anti-predator game or one hawk–dove game. Our individual-based simulations (Fig. 2) confirm the general principle[15] that individuals with higher future expectations (in our case, thorough explorers) should behave in a more cautious way. Superficial explorers evolve to behave boldly (aggressively), whereas thorough explorers evolve low levels of boldness (aggressiveness). This outcome is consistent across replicate simulations and holds for a broad range of parameter conditions (see Supplementary Information).

To show that this principle gives rise to personalities, we next consider scenarios where individuals play several risky games between year 1 and year 2. We stress that, in principle, fully flexible behaviour could evolve in all our simulations. For example, an individual that behaves aggressively in a first hawk–dove game could behave shyly in an anti-predator game and show any level of aggression in a subsequent hawk–dove game. Yet, we find that selection gives rise to stable individual differences within the same game and correlated behavioural traits across different games (Fig. 3). Figure 3a considers the scenario where individuals play two hawk–dove games sequentially. Here superficial explorers evolve high levels of aggression in both games, whereas thorough explorers evolve to be consistently non-aggressive. Figure 3b depicts the scenario where individuals face one anti-predator and one hawk–dove game sequentially. Here evolution gives rise to the coexistence of superficial explorers that are both bold and aggressive and thorough explorers that are shy and non-aggressive. These results are consistent across replicate simulations (Fig. 3c, d), they hold for a broad range of parameter conditions (see Supplementary Information) and they extend to more complex situations where individuals play more than two games (not shown). In other words, we robustly observe the well-known behavioural syndrome linking explorative behaviour, aggression and boldness, which has been reported for numerous species in a diversity of taxa[5,7,8].

Up to now we have assumed, for simplicity, that individuals reproduce asexually. We obtain qualitatively the same results for scenarios that allow for recombination and diploid genetics, as long as individual alleles have a large phenotypic effect. Under this condition disruptive selection gives rise to a small number of discrete phenotypes[25,26], each corresponding to a distinct personality type. A more realistic approach, however, would be to consider quantitative traits, which are often thought to be influenced by many loci with small effects. We now incorporate such quantitative genetics into our model (see Methods).

Consider a scenario where individuals face one anti-predator and one hawk–dove game sequentially. As shown in Fig. 4a, disruptive selection does not now result in two extreme exploration strategies but in the stable coexistence of a broad range of explorative behaviours. Similarly, with respect to both boldness and aggressiveness,

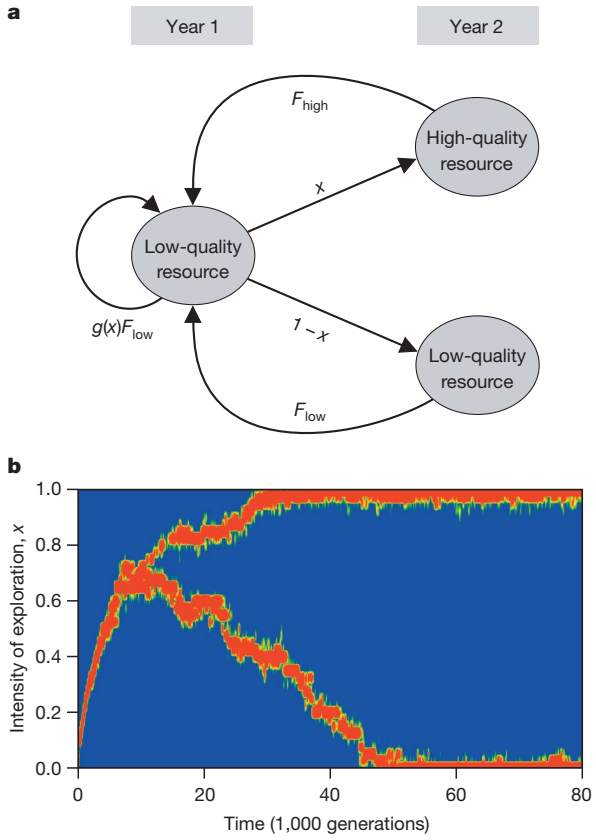


Figure 1 | Stable coexistence of exploration strategies owing to a life-history trade-off. **a**, Structure of the life-history model illustrating the trade-off between current and future reproduction. Natural selection acts on the exploration intensity x that corresponds to the probability of finding a high-quality resource in the future. Although having a positive effect on reproduction in year 2 (future fitness), a high value of x decreases the probability $g(x)$ of obtaining reproductive resources in year 1. F_{low} and F_{high} denote the reproductive output in the case of a low- and high-quality resource, respectively. **b**, The trade-off in **a** induces disruptive selection on exploration intensity and gives rise to the stable coexistence of superficial ($x = 0$) and thorough ($x = 1$) explorers.

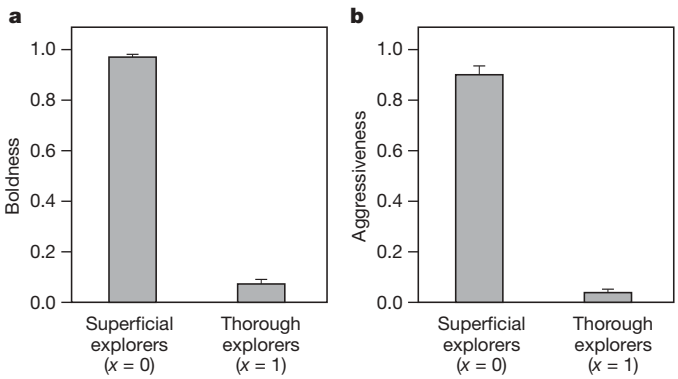


Figure 2 | Evolution of variation in risk-taking behaviour. The outcome of evolution after 3×10^5 generations of selection where individuals following the life cycle depicted in Fig. 1 play a single risky game between year 1 and year 2. **a**, In the case of an anti-predator game superficial explorers evolve high levels of boldness, whereas thorough explorers show low levels of boldness. **b**, In the case of a hawk–dove game superficial explorers are aggressive, whereas thorough explorers are non-aggressive. The bars correspond to mean trait values averaged over ten replicate simulations (error bars indicate standard errors).

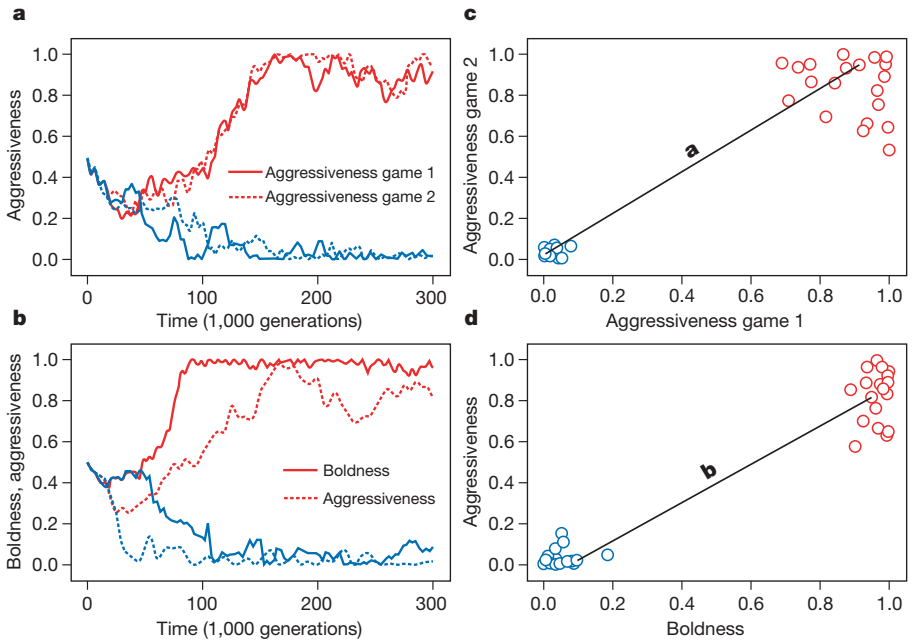


Figure 3 | Evolution of personalities. Simulations illustrating the evolution of consistent individual differences between superficial explorers (red) and thorough explorers (blue). **a**, When individuals face two hawk–dove games superficial explorers evolve high levels of aggressiveness in both games, whereas thorough explorers are consistently non-aggressive. **b**, Confronted with both an anti-predator and a hawk–dove game a behavioural syndrome evolves: superficial explorers are bold and aggressive, whereas thorough explorers are shy and non-aggressive. These outcomes are robust across replicate simulations. **c** and **d** summarize the outcome of 20 replicate simulations for the scenarios in **a** and **b**, respectively. Each simulation is represented by two circles corresponding to the evolved trait combinations of superficial and thorough explorers. Black lines indicate the trait combinations for the simulations depicted in **a** and **b**.

we find a gradation of behavioural traits ranging from low to high levels (Fig. 4b, c). A clear pattern emerges: the more superficially an individual explores its environment (Fig. 4a), the more boldly it behaves in the anti-predator game (Fig. 4b) and the more aggressive it is in the hawk–dove game (Fig. 4c). In other words, in the case of quantitative genetic variation we find a whole spectrum of personality types, which is in line with many empirical studies[12,27] (but see ref. 5).

To sum up, our model offers a plausible explanation for individual differences within a population, the evolution of behavioural correlations within and across contexts and the fact that particular traits such as explorative behaviour, boldness and aggressiveness tend to be associated. Our theory is well testable by predicting (1) under what circumstances behavioural correlations should occur, (2) what particular traits should be correlated and (3) what sign the correlations should take. Behavioural correlations are to be expected whenever individual differences in residual reproductive value (that is, expected future fitness) occur. One might think of the difference between high- and low-ranking individuals in a dominance hierarchy, between dispersers and philopatric individuals in a metapopulation or between residents and floaters in a territorial system. In any such situation, we would expect correlations between those behavioural traits that involve risks that might prevent individuals from reaping the returns from reproductive investments. In addition to intraspecific aggression and boldness, one might also think of behaviours such as brood defence, nest guarding or conspicuous displays to attract mates. All other things being equal, we would expect such traits to be positively correlated.

Our model certainly does not explain all aspects of animal personalities (for example, cooperativeness[28]) and alternative evolutionary mechanisms may also have an important role (refs 10, 11, 29, 30). Yet, as argued above, we believe that our line of argument applies to a broad class of ecological situations. In all these situations the same basic principle will give rise to the evolution of animal personalities: the more an individual has to lose, the more risk-averse it should be, across contexts and through time.

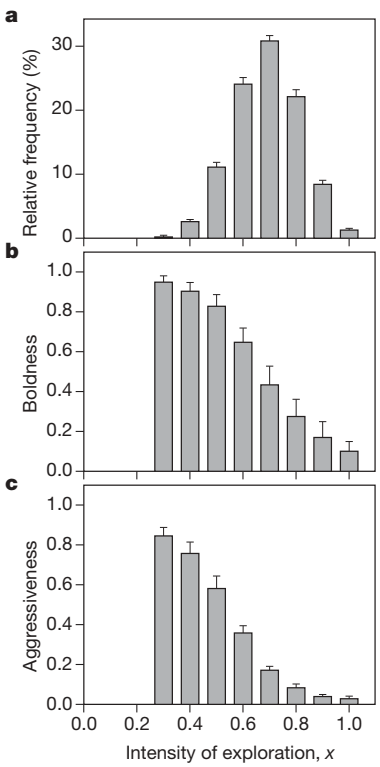


Figure 4 | Evolution of continuous variation in personalities. **a**, If all traits are encoded by multiple loci with small effects, disruptive selection on the exploration intensity x does not result in two discrete phenotype classes but in a broad distribution of exploration strategies. The graphs depict the average level of boldness (**b**) and aggressiveness (**c**) that evolved as a function of x within 2×10^5 generations (mean of ten replicate simulations; error bars indicate standard errors). The evolved levels of boldness and aggressiveness are strongly correlated, corresponding to a boldness–aggressiveness syndrome (correlation coefficients in the ten simulations ranged from 0.40 to 0.81, with a mean value of 0.65).

METHODS SUMMARY

Our conclusions are based on general arguments that are supported by individual-based simulations (main text) and analytical results based on evolutionary invasion considerations (Supplementary Information). We consider a population where individuals follow the basic life cycle illustrated by Fig. 1a. Moreover, between year 1 and year 2, each individual plays one or more anti-predator games and/or one or more hawk–dove games. Individuals are characterized by a suite of heritable traits corresponding to (1) their life-history strategy x, (2) for each anti-predator game the tendency to be bold, and (3) for each hawk–dove game the tendency to be aggressive. Individuals could in principle evolve fully flexible behaviour. For example, an individual that behaves aggressively in a first hawk–dove game could exhibit low levels of aggressions in a subsequent hawk–dove game.

The reproductive success of individuals is frequency- and density-dependent and reflects the fecundity associated with the life-history strategy, the mortality risks associated with bold and aggressive behaviour, and the payoff accumulated in the games. The resulting fitness function is analysed in the Supplementary Information by means of an invasion analysis. In the main text, the assumptions are implemented in individual-based simulations in which trait frequencies change over time under the influence of natural selection. The simulations were run until evolutionary equilibrium was reached. The resulting population was analysed focusing on three key questions. First, does evolution give rise to the coexistence of life-history strategies? Second, does evolution result in stable differences between individuals if the same game is played repeatedly? And third, does evolution lead to behavioural correlations between boldness and aggressiveness?

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 March; accepted 13 April 2007.

1. Pervin, L. A. & John, O. P. (eds) *Handbook of Personality: Theory and Research* (The Guilford Press, New York, 1999).
2. Wilson, D. S., Clark, A. B., Coleman, K. & Dearstyne, T. Shyness and boldness in humans and other animals. *Trends Ecol. Evol.* **9,** 442–446 (1994).
3. Boissy, A. Fear and fearfulness in animals. *Q. Rev. Biol.* **70,** 165–191 (1995).
4. Gosling, S. D. & John, O. P. Personality dimensions in nonhuman animals: a cross-species review. *Curr. Dir. Psychol. Sci.* **8,** 69–75 (1999).
5. Koolhaas, J. M. *et al.* Coping styles in animals: current status in behavior and stress-physiology. *Neurosci. Biobehav. Rev.* **23,** 925–935 (1999).
6. Gosling, S. D. From mice to men: What can we learn about personality from animal research? *Psychol. Bull.* **127,** 45–86 (2001).
7. Sih, A., Bell, A. M., Johnson, J. C. & Ziemba, R. E. Behavioral syndromes: an integrative overview. *Q. Rev. Biol.* **79,** 241–277 (2004).
8. Groothuis, T. G. G. & Carere, C. Avian personalities: characterization and epigenesis. *Neurosci. Biobehav. Rev.* **29,** 137–150 (2005).
9. Coleman, K. & Wilson, D. S. Shyness and boldness in pumpkinseed sunfish: individual differences are context-specific. *Anim. Behav.* **56,** 927–936 (1998).
10. Wilson, D. S. Adaptive individual differences within single populations. *Phil. Trans. R. Soc. Lond. B* **353,** 199–205 (1998).
11. Dall, S. R. X., Houston, A. I. & McNamara, J. M. The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecol. Lett.* **7,** 734–739 (2004).
12. Bell, A. M. Behavioural differences between individuals and two populations of stickleback (*Gasterosteus aculeatus*). *J. Evol. Biol.* **18,** 464–473 (2005).
13. Roff, D. A. *Life History Evolution* (Sinauer, Sunderland, 2002).
14. Rueffler, C., Van Dooren, T. J. M. & Metz, J. A. J. Adaptive walks on changing landscapes: Levins' approach extended. *Theor. Popul. Biol.* **65,** 165–178 (2004).
15. Clark, C. W. Antipredator behavior and the asset-protection principle. *Behav. Ecol.* **5,** 159–170 (1994).
16. Drent, P. J., Van Oers, K. & Van Noordwijk, A. J. Realized heritability of personalities in the great tit (*Parus major*). *Proc. R. Soc. Lond. B* **270,** 45–51 (2003).
17. Dingemanse, N. J., Both, C., Drent, P. J., Van Oers, K. & Van Noordwijk, A. J. Repeatability and heritability of exploratory behaviour in great tits from the wild. *Anim. Behav.* **64,** 929–938 (2002).
18. Bakker, T. C. M. Aggressiveness in sticklebacks (*Gasterosteus aculeatus*)—a behavior–genetic study. *Behaviour* **98,** 1–144 (1986).
19. Sinn, D. L., Apiolaza, L. A. & Moltschaniwskyj, N. A. Heritability and fitness-related consequences of squid personality traits. *J. Evol. Biol.* **19,** 1437–1447 (2006).
20. Sih, A., Bell, A. & Johnson, J. C. Behavioral syndromes: an ecological and evolutionary overview. *Trends Ecol. Evol.* **19,** 372–378 (2004).
21. Reale, D., Gallant, B. Y., Leblanc, M. & Festa-Bianchet, M. Consistency of temperament in bighorn ewes and correlates with behaviour and life history. *Anim. Behav.* **60,** 589–597 (2000).
22. Huntingford, F. A. Relationship between anti-predator behavior and aggression among conspecifics in the three-spined stickleback, *Gasterosteus aculeatus. Anim. Behav.* **24,** 245–260 (1976).
23. Bell, A. M. & Stamps, J. A. Development of behavioural differences between individuals and populations of sticklebacks, *Gasterosteus aculeatus. Anim. Behav.* **68,** 1339–1348 (2004).
24. Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge Univ. Press, Cambridge, 1982).
25. Rueffler, C., Van Dooren, T. J. M., Leimar, O. & Abrams, P. A. Disruptive selection and then what? *Trends Ecol. Evol.* **21,** 238–245 (2006).
26. Van Doorn, G. S. & Dieckmann, U. The long-term evolution of multi-locus traits under frequency-dependent disruptive selection. *Evolution* **60,** 2226–2238 (2006).
27. Dingemanse, N. J., Both, C., Drent, P. J. & Tinbergen, J. M. Fitness consequences of avian personalities in a fluctuating environment. *Proc. R. Soc. Lond. B* **271,** 847–852 (2004).
28. McNamara, J. M., Barta, Z. & Houston, A. I. Variation in behaviour promotes cooperation in the prisoner's dilemma game. *Nature* **428,** 745–748 (2004).
29. Rands, S. A., Cowlishaw, G., Pettifor, R. A., Rowcliffe, J. M. & Johnstone, R. A. Spontaneous emergence of leaders and followers in foraging pairs. *Nature* **423,** 432–434 (2003).
30. Hedrick, A. V. Crickets with extravagant mating songs compensate for predation risk with extra caution. *Proc. R. Soc. Lond. B* **267,** 671–675 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. J. Dingemanse, T. W. Fawcett and I. Pen for discussions and critical reading of the manuscript and D. Visser for preparing the figures.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.J.W. (f.j.weissing@rug.nl).

METHODS

Basic model. We first consider an asexual population of haploid individuals. Each individual is characterized by the allelic values at $k+1$ loci, where one locus determines the exploration behaviour and the other loci determine the behaviour in the k games played between year 1 and year 2. Allelic values range between 0 and 1 and correspond to the exploration intensity x or the strategy in a particular game (probability of being bold or aggressive).

In the first year, an individual with exploration intensity x produces $g(x)F_{\mathrm{low}}$ offspring, where $g(x) = (1-x)^{\beta}$. All figures are based on $\beta = 1.25$, but similar results are obtained for all $\beta > 1$. Between years 1 and 2, an individual accumulates payoffs in the games it participates in, but it also runs the risk of dying (see below). If an individual survives, its reproductive output in year 2 is increased by the payoff it accumulated in these games. All figures shown are based on $f_{\mathrm{high}} = 3.5$, $f_{\mathrm{low}} = 3.0$ and $\alpha = 0.005$, but similar results are obtained as long as $2f_{\mathrm{low}} > f_{\mathrm{high}} > 1$.

During reproduction, mutations occur with a small probability $\mu = 2 \times 10^{-3}$. Mutations have a small effect: they change the allelic value by a value that is drawn from a normal distribution with mean zero and standard deviation 0.02, with the constraint that allelic values remain in the interval from 0 to 1. All results are independent of the specific parameter values for the mutational process and the initial conditions.

Anti-predator games and aggressive encounters. In each of the anti-predator (hawk–dove) games, an individual behaves boldly (aggressively) with a probability determined by its genotype. In an anti-predator game, a bold individual obtains a payoff b but dies with probability γ, whereas a shy individual obtains no payoff and always survives. Aggressive encounters are modified hawk–dove games where individuals fight for a resource of value V. In such an encounter, individuals are paired at random. Payoffs are obtained as in the standard hawk–dove game[24] with one exception: if two aggressive ('hawk') individuals meet, one gets V while the opponent gets 0 and moreover dies with probability δ. All figures are based on $b = 0.1$ and $\gamma = 0.1$ for each of the anti-predator games and $V = 0.1$ and $\delta = 0.5$ for each of the hawk–dove games.

Quantitative genetics. In the model underlying Fig. 4, we consider a sexual population of diploid individuals in which each behaviour is governed by multiple loci with small effects. We consider a scenario with one anti-predator and one hawk–dove game. In total there are five different traits (see below), each of which is governed by a set of five unlinked diallelic loci. There is additive interaction within and across loci, implying that there are in total 11 equidistant phenotypic values for each trait. The first trait corresponds to the exploration strategy, where the different genotypes correspond to 11 exploration tendencies ranging from 0 to 1 in steps of 0.1. The strategy in the anti-predator (hawk–dove) game is modelled as a norm of reaction[13]. The shape of the reaction norm is characterized by two genetically determined parameters a and b such that an individual with exploration intensity x behaves boldly (aggressively) with a probability given by the logistic function $1/\{1 + \exp[-b(x-a)]\}$. The value of a corresponds to the exploration intensity at which both behavioural options (bold–shy and hawk–dove, respectively) are chosen with equal probability, whereas b determines the slope of the reaction norm at $x = a$. As described above, a and b are each encoded by a set of five loci with range restrictions $0 \le a \le 1$ and $-25 \le b \le 25$.

Payoffs are obtained as described above, individuals mate at random, and the number of offspring produced per individual at the end of each year is proportional to the total payoff obtained in that year. With a small probability ($\mu = 2 \times 10^{-5}$) a mutation occurs at a randomly chosen locus. When this happens, the affected allele changes into the alternative allele.

LETTERS

Incipient speciation by divergent adaptation and antagonistic epistasis in yeast

Jeremy R. Dettman[1], Caroline Sirjusingh[1], Linda M. Kohn[1] & James B. Anderson[1]

Establishing the conditions that promote the evolution of reproductive isolation and speciation has long been a goal in evolutionary biology[1–3]. In ecological speciation, reproductive isolation between populations evolves as a by-product of divergent selection and the resulting environment-specific adaptations[4–6]. The leading genetic model of reproductive isolation predicts that hybrid inferiority is caused by antagonistic epistasis between incompatible alleles at interacting loci[1,7]. The fundamental link between divergent adaptation and reproductive isolation through genetic incompatibilities has been predicted[1,4,5], but has not been directly demonstrated experimentally. Here we empirically tested key predictions of speciation theory by evolving the initial stages of speciation in experimental populations of the yeast *Saccharomyces cerevisiae*. After replicate populations adapted to two divergent environments, we consistently observed the evolution of two forms of postzygotic isolation in hybrids: reduced rate of mitotic reproduction and reduced efficiency of meiotic reproduction. This divergent selection resulted in greater reproductive isolation than parallel selection, as predicted by the ecological speciation theory. Our experimental system allowed controlled comparison of the relative importance of ecological and genetic isolation, and we demonstrated that hybrid inferiority can be ecological and/or genetic in basis. Overall, our results show that adaptation to divergent environments promotes the evolution of reproductive isolation through antagonistic epistasis, providing evidence of a plausible common avenue to speciation and adaptive radiation in nature.

The development of reproductive isolation is a key aspect of speciation because it is important for both initial divergence and maintenance of distinct species. If two species are specialized to different environments, interspecific hybrids might have reduced fitness in both environments because they exhibit maladaptive intermediate phenotypes, underscoring the inherent link between ecological isolation and divergent adaptation[5,6,8,9]. The link between genetic isolation and divergent adaptation can be explained by Dobzhansky–Muller genic incompatibilities[1,7]. A progenitor population enters and adapts to divergent environments, and in each of these environments certain new alleles are favoured by selection. Hybridization creates novel combinations of alleles untested by selection[10,11], and negative interactions among such alleles can reduce hybrid fitness.

We investigated the relationship between adaptation to divergent environments and evolution of reproductive isolation using large replicate experimental populations of the yeast *Saccharomyces cerevisiae*. Twelve populations were initiated from a single diploid progenitor (P) and evolved for 500 generations by serial transfer in one of two sub-optimal liquid environments: high-salinity (S) or low-glucose minimal medium (M). Each of the 12 evolved populations displayed a significant increase in relative fitness (>1.0; one-tailed, $t > 3.15$, degrees of freedom d.f. $= 2$, $P < 0.05$ in each case), providing clear evidence for adaptation (Fig. 1). When competing in their

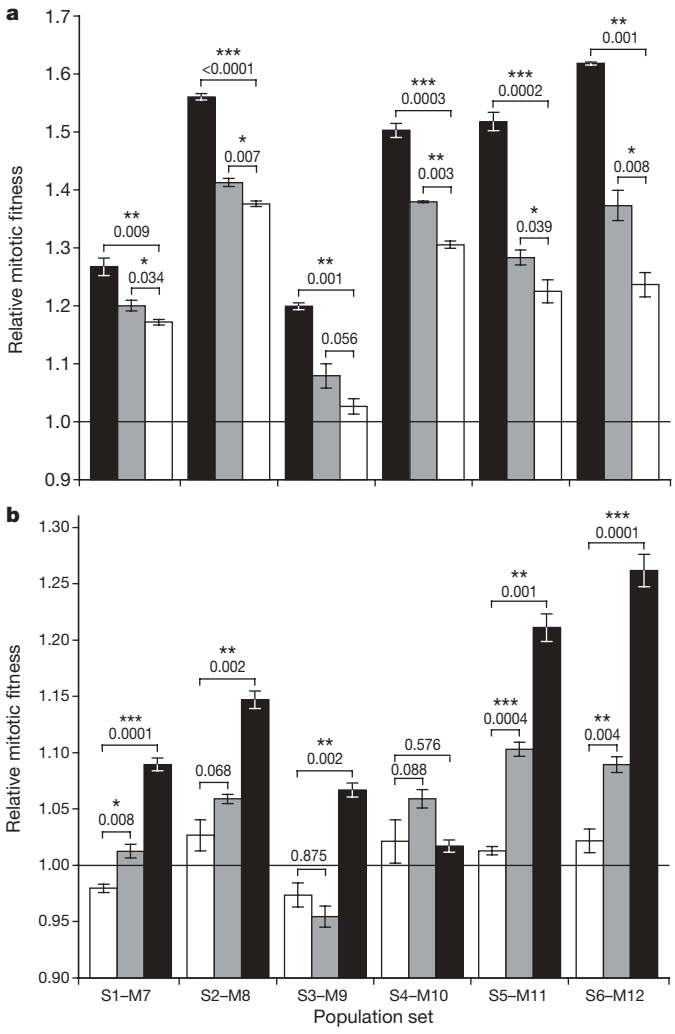


Figure 1 | Comparisons of mitotic fitness. a, Fitness assays in the high-salinity environment. Mean relative fitness estimates for populations evolved in high salinity (S, black), for hybrids between evolved populations and progenitors (S/P, grey), and for hybrids between populations evolved in high salinity and populations evolved in low glucose (S/M, white). **b**, Fitness assays in the low-glucose environment. Mean relative fitness estimates for populations evolved in low glucose (M, black), for hybrids between evolved populations and progenitors (M/P, grey), and for hybrids between populations evolved in high salinity and populations evolved in low glucose (S/M, white). Statistical significance (P values) of comparisons involving S/M hybrids is shown above the brackets. Error bars, ± 1 s.e.m. ($n = 3$). $^{*}P < 0.05$; $^{**}P < 0.005$; $^{***}P < 0.0005$.

[1]Department of Ecology & Evolutionary Biology, University of Toronto, Mississauga, Ontario, L5L 1C6, Canada.

selective environment, evolved populations had an average fitness advantage of 28.8% over the progenitor. By contrast, when competing in the permissive or alternative selective environment, the mean relative fitness of evolved populations was not significantly increased (one-tailed, $t < 1.58$, d.f. $= 11$, $P > 0.42$ in each case). The fitness advantages gained during evolution represented environment-specific adaptations rather than an overall increase in general fitness (Fig. 2).

The hallmark of ecological speciation is environment-dependent (conditional) isolation, in which the fitness of hybrids is lower than that of pure populations in their selective environments. Each S population was paired with an M population, creating six independent comparisons. We tested for reproductive isolation by measuring fitness of the S/M hybrid diploids in all environments (permissive, high salinity and low glucose). In high salinity, S/M hybrid fitness was significantly lower than the S population fitness for each of the six population sets (Fig. 1). Similarly, in low glucose, S/M hybrid fitness was significantly lower than the M population fitness in five of six sets (Fig. 1). The same patterns of hybrid inferiority were observed when data were pooled ($P \leq 0.0001$ for each environment). Conversely, competition in the permissive environment revealed no significant difference between the mean fitness of S/M hybrids and that of pure populations (one-tailed, $t = 0.09$, d.f. $= 36$, $P = 0.53$), indicating that hybrid fitness reductions were dependent on the selective environments. As a control, we also constructed hybrids between populations that had evolved in the same environment. Overall, the fitness of hybrids between parallel-adapted populations was not significantly lower than that of pure populations (one-tailed, $P \geq 0.35$ for each environment). These patterns of conditional reproductive isolation can result only from environment-based divergent selection[12,13].

Our experimental yeast system provided the unique ability to construct hybrids between evolved populations (S or M) and the unevolved progenitor (P), resulting in 'progenitor hybrids' (S/P and M/P). Fitness gains of progenitor hybrids, which possessed only one evolved genome component, were only 0.51 ± 0.08 (mean $\pm$ standard error of mean, s.e.m.) of that of pure populations, which possessed two evolved genome components. Thus, reduced hybrid fitness was caused in part by incomplete dominance of the net effects of beneficial adaptive factors. These data provide a clear example of ecological isolation, because hybrid inferiority is determined by an interaction between genotype and environment[12,13].

Progenitor hybrids and S/M hybrids should have equal fitness if fitness is determined solely by adaptive mutations. In high salinity, S/M hybrids were less fit than S/P hybrids in all six population sets, and significantly so for five sets (Fig. 1a). In low glucose, S/M hybrids were less fit than M/P hybrids in five sets, and significantly so for three sets (Fig. 1b). The inferiority of S/M hybrids compared to progenitor hybrids can be explained by two factors: first, antagonistic epistasis[10,14] between S and M genome components, and, second, detrimental pleiotropic effects of adaptation on fitness in the alternative environment. The percentage of total reduction in S/M hybrid fitness caused by antagonistic epistasis and pleiotropy together averaged $27.6 \pm 3.5\%$. In general, the detrimental pleiotropic effects of adaptation were relatively minor (Fig. 2). Mean fitness for the 12 pure populations in the alternative environment was 0.98 ± 0.01, suggesting that antagonistic epistasis was responsible for a considerable proportion of S/M hybrid fitness deficits. For the five comparisons in which detrimental effects of pleiotropy were absent (fitness of pure population in alternative environment ≥ 1.0; S1–M7 in high salinity, and S1–M7, S2–M8, S4–M10 and S6–M12 in low glucose (see Methods)), S/M hybrids were significantly less fit than progenitor hybrids (paired-sample t-test, $P < 0.0001$; see Fig. 1 for individual comparisons). In these cases, S/M hybrid inferiority was caused solely by genetic isolation by means of antagonistic epistasis.

In addition to mitotic reproduction, S/M hybrids were severely inferior to pure populations in their ability to reproduce meiotically (sexually). In an environment permissive to sexual reproduction, mean meiotic efficiency of pure populations was $63.6 \pm 1.2\%$, whereas S/M hybrids had a mean of only $34.4 \pm 1.7\%$. For all six independent comparisons, meiotic efficiency of S/M hybrids was significantly lower than that of both pure populations (Fig. 3). The mean meiotic efficiency of hybrids between parallel-adapted populations (45.84 ± 1.58) was significantly greater than the S/M hybrid mean (one-tailed, $t = 4.97$, d.f. $= 58$, $P < 0.0001$), but was still less than that of the pure populations (one-tailed, $t = 8.89$, d.f. $= 64$, $P < 0.0001$). Selection in divergent environments resulted in greater reproductive isolation than parallel selection in the same environment. These unconditional reductions in hybrid meiotic efficiency represent another clear example of genetic isolation by means of antagonistic epistasis between divergently adapted genomic components.

We are able to reject the two alternative explanations for reduced meiotic reproduction of hybrids: chromosomal rearrangements[15–17] and the anti-recombination activity of the mismatch repair system[18,19]. Both mechanisms have extremely detrimental effects on progeny viability resulting from abnormal disjunction, aneuploidy and deficiency[16–19]. In contrast, meiotic progeny from S/M hybrids all had high viabilities, ranging from 97.7–100% (mean $= 99.4\% \pm 0.4$), with normal mendelian segregation of markers. This observation

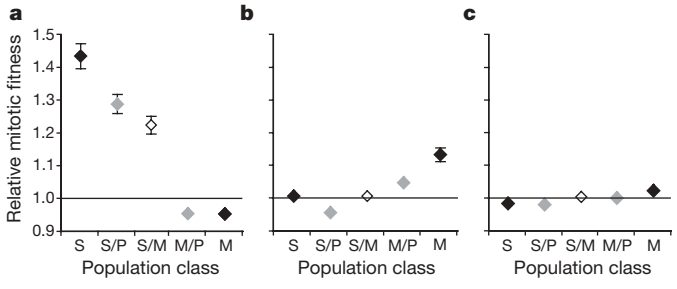


Figure 2 | General patterns of mitotic fitness across population classes. Mean relative fitness estimates for pure populations (S and M, black filled), progenitor hybrids (S/P and M/P, grey filled), and S/M hybrids (S/M, white filled) assayed in: **a**, the high-salinity environment, **b**, the low-glucose environment, and **c**, the permissive environment. Data were pooled for all six populations sets; error bars, ± 1 s.e.m. ($n = 18$), if greater than the symbol size.

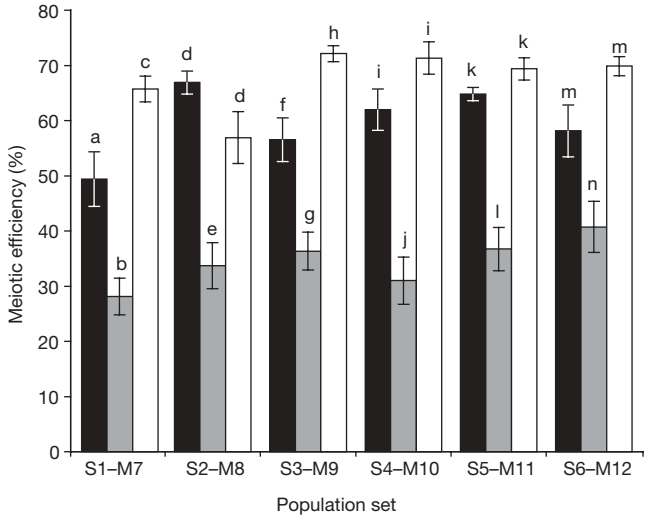


Figure 3 | Comparisons of meiotic efficiency. Mean meiotic efficiency values for pure populations (S, black; M, white) and for S/M hybrids (S/M, grey) from six population sets. Error bars, ± 1 s.e.m. ($n = 5$). For all six sets, the meiotic efficiency of hybrids was significantly less than that of both pure populations. Means that are not connected by the same letter are significantly different from each other, as determined by analysis of variance (all $F \geq 13.94$, $P \leq 0.0007$) and post-hoc Tukey tests ($P = 0.05$) performed independently for each population set.

alone excludes both alternative mechanisms (see Supplementary Information 1 for additional evidence).

To investigate whether mitotic fitness and meiotic efficiency of hybrids were affected by interactions between the same genetic factors, we compared these 2 phenotypes across 12 series of 6 hybrid diploids that had variable numbers of S-adapted and M-adapted factors (see Supplementary Methods and Supplementary Information 2). The S haploid component was fixed within each series, so differences in phenotype are caused by the variable number of M-adapted factors. A positive correlation between meiotic efficiency and mitotic fitness (in high salinity), which was predicted if the two phenotypes were affected by the same genetic interactions, was displayed by 5 of the 12 hybrid series, only one of which was significant ($R^2 = 0.78$, $P < 0.02$). Reductions in the two phenotypes were not inextricably linked, which suggests that a complex array of multiple independent interactions cumulatively resulted in concomitant negative effects on both phenotypes in S/M hybrids (Figs 1–3).

Microarray experiments were performed to compare genome-wide gene expression during sporulation[20] between non-hybrid and hybrid states of S and M genome components (see Supplementary Methods and Supplementary Information 3). Reference RNA was an equal mixture of RNA from two pure populations, and test RNA was from S/M hybrids. For the two population sets examined (S2–M8 and S6–M12), a consensus set of four genes displayed expression that was significantly altered in S/M hybrids: two genes were consistently underexpressed and two genes were overexpressed. Both underexpressed genes (*RCR1* and YLR194C) are involved in cell wall synthesis and organization; their mis-expression might affect both mitotic fitness and meiotic efficiency. An additional 31 genes that had altered expression were specific to one of the two population sets examined. Two sporulation genes (*SPS100* and *PFS1*) were underexpressed, providing evidence for disruption of the sporulation transcriptional cascade in hybrids and an entry point for future investigation of the underlying mechanism(s) of antagonistic epistasis.

Previous studies found no evidence for recessive[21] or dominant[22] incompatibilities causing yeast hybrid sterility. These studies differed from ours in one fundamental way: we examined *de novo* genetic isolation between recently diverged populations rather than pre-established isolation between species that diverged ~5 million years ago. The history of established species may comprise multiple episodes of adaptation, with successive gains or losses of mechanisms of isolation. The initial mechanisms of isolation are obscure.

Unlike natural species, our experimental populations have an evolutionary history that is known with certainty. We can therefore conclude that divergent adaptation caused the reproductive isolation observed in this investigation. Experimental evolution of reproductive isolation has been studied in a few eukaryotes (mainly *Drosophila*) with mixed results[23–27]. Previous research has focused mostly on prezygotic isolation, and we are aware of only a single study that reported successful evolution of postzygotic isolation by means of divergent selection[28]. We present the most striking example of experimental evolution of postzygotic isolation observed in any organism, and the first for the fungal kingdom[29].

Although the isolation that evolved *de novo* in our short-term experiment is partial, it represents incipient speciation. Given more time, complete reproductive isolation is likely to evolve. Effects of reproductive isolation are compounded over successive steps of the life cycle; pure populations have a 2.14-fold advantage (range, 2.00–2.31) over S/M hybrids when mitotic and meiotic reproduction are combined. Furthermore, these effects are multiplied over successive generations, allowing minor hybrid inferiority to have major detrimental consequences to overall gene flow among natural populations. Therefore, the most significant point in the speciation process is the initial development of reproductive isolation, regardless of its intensity, as studied here.

Our results validate the theoretical link between divergent adaptation and reproductive isolation. We show that specialization to alternative environments provides the phenotypic variation on which ecological mechanisms of reproductive isolation can operate. We also show that the processes of mutation and selection during divergent adaptation can facilitate the accumulation of antagonistically interacting genetic factors. Overall, our experiments support the theory that divergent adaptation promotes the evolution of reproductive isolation, caused by intrinsic genetic incompatibilities and/or extrinsic environmental factors, and is a major driving force for speciation.

METHODS SUMMARY

In the diploid progenitor strains, the entire *URA3* open reading frame was replaced with one of two antibiotic resistance cassettes (geneticin or nourseothricin) flanked by one of five DNA barcodes. Each of the 12 experimental populations was propagated in 10 ml of shaking liquid medium, with daily serial transfers (1/100 dilution) for a total of 500 generations. Six populations evolved in rich high-salinity medium (S1–S6), and six populations evolved in low-glucose minimal medium (M7–M12). Six independent population sets were designated by pairing S and M populations with alternative antibiotic resistance. To create hybrid populations, pure populations were sporulated, mated with each other, and then grown in medium containing both antibiotics to select for hybrids. The same procedure was used to create hybrids between pure populations and progenitors.

Mitotic fitness of populations was determined by direct competition against an alternatively tagged progenitor and by measurement of change in proportions of strains in mixed culture over time[30]. Genomic DNA was extracted from mixed cultures at the start and finish of a 24 h competition period. A segment of the tag region was amplified using polymerase chain reaction and transferred to membranes, which were probed with oligonucleotides complementary to the barcode tags and an invariant region shared by all competitors. The number of cell doublings was calculated from tag signal ratios and optical densities of cultures. Relative fitness was defined as the ratio of cell doublings by the assay population to cell doublings by the progenitor.

The meiotic efficiency of populations was defined as the percentage of diploid cells that underwent meiosis. After 72 h of sporulation, meiotic efficiency was determined by direct counting under ×400 magnification.

Microarray protocols were performed as described in ref. 14 except that cell populations were incubated to mid-log growth phase then transferred to sporulation medium for 12 h before RNA extraction.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 November 2006; accepted 12 April 2007.

1. Dobzhansky, T. *Genetics and the Origin of Species* (Columbia Univ. Press, New York, 1937).
2. Coyne, J. A. Genetics and speciation. *Nature* **355,** 511–515 (1992).
3. Coyne, J. A. & Orr, H. A. *Speciation* (Sinauer, Sunderland, 2004).
4. Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, Oxford, 2000).
5. Schluter, D. Ecology and the origin of species. *Trends Ecol. Evol.* **16,** 372–380 (2001).
6. Rundle, H. D. & Nosil, P. Ecological speciation. *Ecol. Lett.* **8,** 336–352 (2005).
7. Muller, H. J. Isolating mechanisms, evolution, and temperature. *Biol. Symp.* **6,** 71–125 (1942).
8. McKinnon, J. S. *et al.* Evidence for ecology's role in speciation. *Nature* **429,** 294–298 (2004).
9. Funk, D. J., Nosil, P. & Etges, W. J. Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proc. Natl Acad. Sci. USA* **103,** 3209–3213 (2006).
10. Orr, H. A. & Turelli, M. The evolution of postzygotic isolation: accumulating Dobzhansky–Muller incompatibilities. *Evolution* **55,** 1085–1094 (2001).
11. Presgraves, D. C., Balagopalan, L., Abmayr, S. M. & Orr, H. A. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature* **423,** 715–719 (2003).
12. Rundle, H. D. & Whitlock, M. C. A genetic interpretation of ecologically dependent isolation. *Evolution* **55,** 198–201 (2001).
13. Demuth, J. P. & Wade, M. J. On the theoretical and empirical framework for studying genetic interactions within and among species. *Am. Nat.* **165,** 524–536 (2005).
14. Anderson, J. B., Ricker, N. & Sirjusingh, C. Antagonism between two mechanisms of antifungal drug resistance. *Eukaryot. Cell* **5,** 1243–1251 (2006).
15. Rieseberg, L. H. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16,** 351–358 (2001).

16. Delneri, D. *et al.* Engineering evolution to study speciation in yeasts. *Nature* **422,** 68–72 (2003).
17. Fischer, G., James, S. A., Roberts, I. N., Oliver, S. G. & Louis, E. J. Chromosomal evolution in *Saccharomyces*. *Nature* **405,** 451–454 (2000).
18. Hunter, N., Chambers, S. R., Louis, E. J. & Borts, R. H. The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *EMBO J.* **15,** 1726–1733 (1996).
19. Greig, D., Travisano, M., Louis, E. J. & Borts, R. H. A role for the mismatch repair system during incipient speciation in *Saccharomyces*. *J. Evol. Biol.* **16,** 429–437 (2003).
20. Chu, S. *et al.* The transcriptional program of sporulation in budding yeast. *Science* **282,** 699–705 (1998).
21. Greig, D. A screen for recessive speciation genes expressed in the gametes of F1 hybrid yeast. *PLoS Genet.* **3,** e21 (2007).
22. Greig, D., Borts, R. H., Louis, E. J. & Travisano, M. Epistasis and hybrid sterility in *Saccharomyces*. *Proc. R. Soc. Lond. B* **269,** 1167–1171 (2002).
23. Rice, W. R. & Hostert, E. E. Laboratory experiments on speciation: what have we learned in 40 years? *Evolution* **47,** 1637–1653 (1993).
24. Mooers, A. Ø., Rundle, H. D. & Whitlock, M. C. The effects of selection and bottlenecks on male mating success in peripheral isolates. *Am. Nat.* **153,** 437–444 (1999).
25. Rundle, H. D. Divergent environments and population bottlenecks fail to generate premating isolation in *Drosophila pseudoobscura*. *Evolution* **57,** 2557–2565 (2003).
26. Rundle, H. D., Chenoweth, S. F., Doughty, P. & Blows, M. W. Divergent selection and the evolution of signal traits and mating preferences. *PLoS Biol.* **3,** e368 (2005).
27. Leu, J. Y. & Murray, A. W. Experimental evolution of mating discrimination in budding yeast. *Curr. Biol.* **16,** 280–286 (2006).
28. de Oliveira, A. K. & Cordeiro, A. R. Adaptation of *Drosophila willistoni* experimental populations to extreme pH medium II. Development of incipient reproductive isolation. *Heredity* **44,** 123–130 (1980).
29. Kohn, L. M. Mechanisms of fungal speciation. *Annu. Rev. Phytopathol.* **43,** 279–308 (2005).
30. Anderson, J. B. *et al.* Mode of selection and experimental evolution of antifungal drug resistance in *Saccharomyces cerevisiae*. *Genetics* **163,** 1287–1298 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by Discovery grants to J.B.A. and L.M.K., and a Postdoctoral Fellowship to J.R.D., from the Natural Science and Engineering Research Council of Canada.

Author Contributions The research was conceived and planned by all authors. C.S. and J.R.D. performed the experiments, and J.R.D. analysed the data. J.R.D, J.B.A. and L.M.K. contributed to the writing of the manuscript, which was coordinated by J.R.D.

Author Information The full microarray data set has been deposited in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under accession series GSE6870. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.R.D. (jdettman@utm.utoronto.ca).

METHODS

Additional details are given in Supplementary Methods.

Strains. Progenitor strains were derived from diploid strains constructed in ref. 30. The entire *URA3* open reading frame was replaced with a geneticin (G418) antibiotic resistance cassette flanked by two DNA barcodes. The G418 cassette was replaced with a nourseothricin (NAT) resistance cassette in each strain to give ten different combinations of two antibiotic resistance cassettes and five barcode tags.

Evolution regimen. Six populations, designated S1 to S6, were evolved in rich high-salinity medium (environment S, yeast peptone dextrose (YPD: 1% yeast extract, 2% peptone, 2% dextrose) plus 1.0 M NaCl). The other six populations, designated M7 to M12, were evolved in low-glucose minimal medium (environment M, 0.67% yeast nitrogen base without amino acids, 0.25% dextrose, 0.02 mg ml^{-1} uracil). Populations were propagated in 10 ml of liquid medium at 30 °C on a rotary shaker at 250 r.p.m. After each 24 h of growth, which represented an average of 6.64 generations, 100 µl of culture was transferred to 9.9 ml of fresh medium. This cycle was repeated until populations had evolved for a total of 500 generations. Every 100 generations, 8 populations (S1 to S4, M7 to M10) were sporulated, diploidized by intra-population mating, and then used to re-initiate the evolution regimen. Every 100 generations, aliquots of all populations were archived at –80 °C. No cross-contamination was detected, and each tag and resistance marker appeared where expected without exception.

Population sets and construction of hybrids. Population sets were designated by pairing populations that had alternative antibiotic resistance cassettes. To create hybrid populations, the two pure populations were sporulated, mated with each other, and then grown in liquid YPD with both NAT and G418 antibiotics. The same procedure was used to create hybrids between pure populations and alternatively tagged progenitors. All sporulations and matings were performed *en masse*, so assay results represent population averages rather than specific values for single genotypes.

Mitotic fitness. Mitotic fitness of populations was determined by direct competition against an alternatively tagged progenitor and by measurement of change in proportions of strains in mixed culture over time[30]. The two competitors were mixed in equal proportions and competition was initiated (time 0) by inoculating 10 µl into 9.99 ml of competition medium. Competitions were run for 24 h (time 1) under conditions identical to both selective environments (high salinity or low glucose) as well as in a rich permissive environment (YPD). Genomic DNA was extracted from mixed cultures at time 0 and time 1, and a segment of the *URA3* replacement region was amplified by polymerase chain reaction. Amplicons were transferred to membranes, which were probed in succession using oligonucleotides complementary to the barcode tags and an invariant region shared by all competitors. The number of cell doublings was calculated as $\log_2([R_f \times D_f] / [R_i \times D_i])$, where R_f is the ratio of tag signal to invariant probe signal at time 1 and R_i is the corresponding ratio at time 0. D_f and D_i are optical densities of cultures at time 1 and time 0, respectively. Relative fitness was defined as the ratio of cell doublings by the assay population to cell doublings by the progenitor. To estimate the mean dominance of the net effects of all adaptive factors present in the adapted genome, we compared fitness of pure populations to that of progenitor hybrids. Here, $w_{X(Y)}$ equals the relative fitness of population X in the Y environment. The degree of dominance for S-adapted factors was calculated as $(w_{S/P(S)} - 1) / (w_{S(S)} - 1)$, and for M-adapted factors as $(w_{M/P(M)} - 1) / (w_{M(M)} - 1)$. The combined effects of antagonistic epistasis and pleiotropy were estimated by the fitness difference between S/M hybrids and progenitor hybrids. By controlling one haploid component in the diploid, we could determine the effects of the other classes of haploid components, and associated interactions. The proportion of total reduction in S/M hybrid fitness attributable to antagonistic epistasis and pleiotropy together was calculated by $(w_{S/P(S)} - w_{S/M(S)}) / (w_{S(S)} - w_{S/M(S)})$ and $(w_{M/P(M)} - w_{S/M(M)}) / (w_{M(M)} - w_{S/M(M)})$.

Meiotic efficiency. After 72 h of sporulation, the percentage of diploid cells that underwent, or were undergoing, meiosis was determined by direct counting under ×400 magnification. The presence of dyads, triads and tetrads constituted evidence for meiosis. An average of >100 cells were examined per meiotic efficiency replicate.

Microarrays. Protocols were performed as described[14] with the exception that cell populations were incubated to mid-log growth phase and then transferred to the sporulation medium for 12 h before RNA extraction. For each hybrid versus non-hybrid comparison, four independently replicated microarray hybridizations, with balanced dye swapping, were performed. Gene lists were made from genes that were differentially expressed by ≥1.5-fold in S/M hybrid populations relative to the reference.

LETTERS

Symbolic arithmetic knowledge without instruction

Camilla K. Gilmore[1], Shannon E. McCarthy[2] & Elizabeth S. Spelke[2]

Symbolic arithmetic is fundamental to science, technology and economics, but its acquisition by children typically requires years of effort, instruction and drill[1,2]. When adults perform mental arithmetic, they activate nonsymbolic, approximate number representations[3,4], and their performance suffers if this nonsymbolic system is impaired[5]. Nonsymbolic number representations also allow adults, children, and even infants to add or subtract pairs of dot arrays and to compare the resulting sum or difference to a third array, provided that only approximate accuracy is required[6–10]. Here we report that young children, who have mastered verbal counting and are on the threshold of arithmetic instruction, can build on their nonsymbolic number system to perform symbolic addition and subtraction[11–15]. Children across a broad socio-economic spectrum solved symbolic problems involving approximate addition or subtraction of large numbers, both in a laboratory test and in a school setting. Aspects of symbolic arithmetic therefore lie within the reach of children who have learned no algorithms for manipulating numerical symbols. Our findings help to delimit the sources of children's difficulties learning symbolic arithmetic, and they suggest ways to enhance children's engagement with formal mathematics.

We presented children with approximate symbolic arithmetic problems in a format that parallels previous tests of non-symbolic arithmetic in preschool children[8,9]. In the first experiment, five- to six-year-old children were given problems such as "If you had twenty-four stickers and I gave you twenty-seven more, would you have more or less than thirty-five stickers?". Children performed well above chance (65.0%, $t_{19} = 2.77$, $P = 0.012$) without resorting to guessing or comparison strategies that could serve as alternatives to arithmetic. Children who have been taught no symbolic arithmetic therefore have some ability to perform symbolic addition problems.

The children's performance nevertheless fell short of performance on non-symbolic arithmetic tasks using equivalent addition problems with numbers presented as arrays of dots and with the addition operation conveyed by successive motions of the dots into a box (71.3% correct, $F_{1,34} = 4.26$, $P = 0.047$)[8]. Some children appeared to be confused by the wording of the problems; earlier experiments using even more complex wording had yielded chance performance in children of this age[8]. Accordingly, a second experiment tested children's understanding of symbolic addition with simpler questions accompanied by visual displays (Fig. 1a). Children answered questions of the form "Sarah has fifteen candies and she gets nineteen more, John has fifty-one candies. Who has more?" Children's performance was as accurate (73.3%, $t_{19} = 6.40$, $P = 4 \times 10^{-6}$) as the performance of children in past research given such problems in nonsymbolic form[8] (Fig. 2a), and again did not depend on guessing strategies involving comparisons between pairs of arrays.

The above experiments were conducted in a quiet laboratory setting and focused on children living in a relatively wealthy and highly educated community. To investigate their generality, we used the method of the preceding experiment to assess the symbolic arithmetic performance of children from a low- to middle-socioeconomic community, tested in their kindergarten classrooms. These five- to six-year-old children also performed well above chance (63.9%, $t_{36} = 6.273$, $P = 3 \times 10^{-7}$) and without resort to guessing strategies, although performance was lower overall when children were tested in the more distracting classroom environment ($F_{1,55} = 5.56$, $P = 0.022$, Fig. 2b). The children's performance of approximate symbolic arithmetic in this experiment was correlated with their performance on a test of mastery of the school's curriculum, although the specificity and the causal nature of this relationship remain to be established. Capacities for symbolic arithmetic are evidently available to children across a range of social and economic backgrounds, and they can be elicited in classroom as well as laboratory settings.

What processes underlie children's symbolic approximate arithmetic? We first considered whether children might have learned some symbolic addition facts and used their fact knowledge to solve the present problems. However, children failed to provide the exact solution to the present problems, to solve those problems when the comparison quantity was altered so as to require an exact representation of number, or to solve the problems through rounding strategies allowing for single-digit addition or comparison. Children's approximate arithmetic performance evidently does not depend on knowledge of exact number.

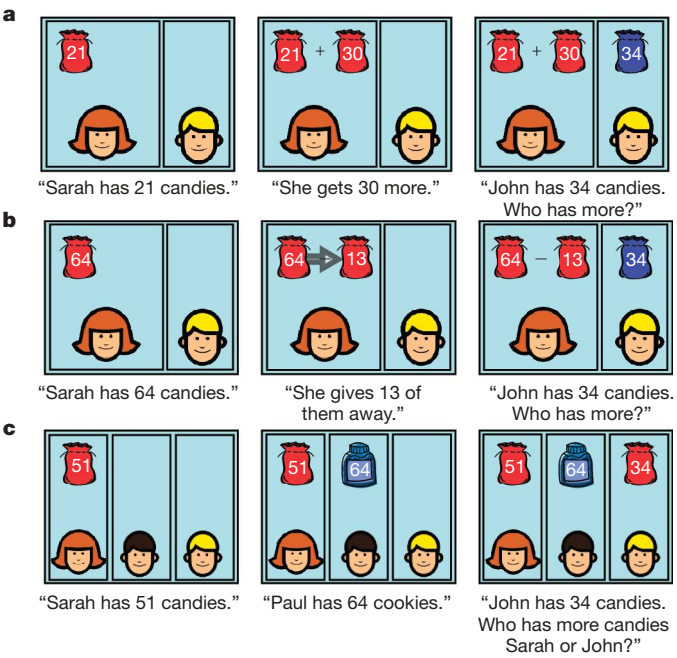


Figure 1 | Example problems of symbolic, approximate arithmetic. **a**, Addition; **b**, subtraction; and **c**, comparison.

[1]Learning Sciences Research Institute, University of Nottingham, Wollaton Road, Nottingham NG8 1BB, UK. [2]Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, Massachusetts 02138, USA.

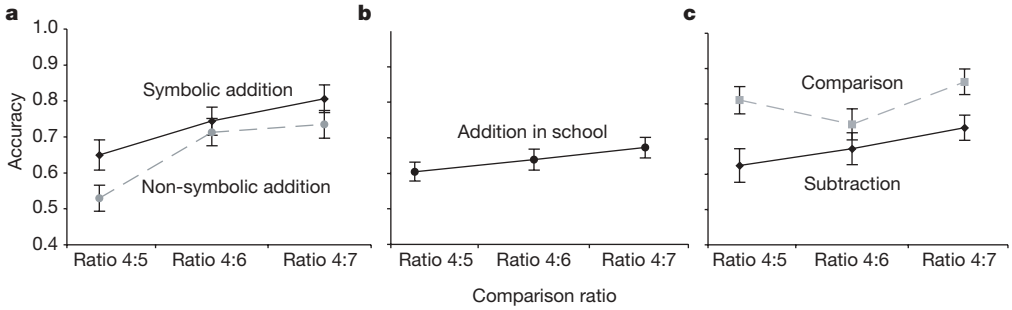


Figure 2 | Five-year-old children's performance on approximate addition, subtraction and comparison problems. a, Children's performance of addition problems in which the sum differed from the comparison number at three ratios. Solid line indicates performance in the present experiment with numbers presented as words and symbols ($n = 20$); dotted line indicates performance in a previous study with the same numerical values presented nonsymbolically as dot arrays[8]. **b**, Children's performance of the same symbolic addition problems, administered in a classroom setting ($n = 37$). **c**, Children's performance of equivalent symbolic subtraction and comparison problems ($n = 29$). Error bars represent s.e.m.

We therefore considered whether children draw on nonsymbolic, approximate number representations in solving these symbolic problems. Nonsymbolic number representations show three signature properties in human adults, children, infants, and non-human animals. First, comparison, addition and subtraction are subject to a ratio limit: accuracy falls as the ratio of the numbers to be compared approaches one[4,8,9,16]. Second, addition is as accurate as comparison: when adults or children sum two arrays and compare the result to a third array, they perform as accurately as when they compare the third array to a single array with the numerosity of the sum[9]. Third, subtraction is less accurate than comparison[9]. If children use nonsymbolic number representations to perform symbolic arithmetic, their performance should show the same signatures.

To test for the ratio signature, we compared children's performance at each of the three ratios. Children's accuracy decreased as the ratio of the sum of the addends to the comparison set approached one (for laboratory-tested children, $F_{1,19} = 14.92$, $P = 2 \times 10^{-5}$, Fig. 2a; for school-tested children, $F_{1,36} = 5.32$, $P = 0.027$, Fig. 2b). The effect of ratio was the same for these symbolic problems as for problems presented non-symbolically[8], and it did not depend on the differential use or effectiveness of strategies that serve as alternatives to addition.

To test for the other two signatures, we presented a new group of children from the same population as those in the previous laboratory experiments with symbolic subtraction and comparison tasks (Fig. 1b, c). Children solved subtraction problems (67.7% correct, $t_{19} = 5.55$, $P = 7 \times 10^{-6}$; Fig. 2c) without basing their answers on exact number knowledge or on a range of alternative strategies and with a ratio limit on subtraction performance. Performance on the comparison problems (80.4% correct, $t_{19} = 9.82$, $P = 1 \times 10^{-10}$) did not differ from performance on the symbolic addition problems ($F_{1,47} = 2.20$, $P = 0.145$) but was higher than performance on the symbolic subtraction problems ($F_{1,27} = 16.58$, $P = 0.0004$). Nevertheless, a direct comparison of performance on the symbolic addition and subtraction problems revealed no significant advantage for addition ($F_{1,46} = 1.36$, $P = 0.250$).

We have shown that children with no formal arithmetic instruction perform symbolic addition and subtraction with limits that reveal three signatures of the non-symbolic arithmetic system: a ratio effect on accuracy, addition performance that is as accurate as direct comparison, and subtraction performance that is less accurate than comparison. These common signatures suggest that children recruit their nonsymbolic number knowledge when they confront new problems of approximate symbolic arithmetic. Once children have learned the verbal counting system and mapped number words onto nonsymbolic representations of number, they spontaneously use nonsymbolic system processes to manipulate quantities presented symbolically. Recruitment of the nonsymbolic system therefore does not depend on instruction in arithmetic.

Arithmetic instruction in most primary school curricula focuses primarily or exclusively on exact operations on small numbers. Most children take years to master the set of exact, single-digit addition and multiplication facts[17], and mastery of these facts is vulnerable to fatigue or interference even in adults[18]. For this reason, the teachers in our school-based study were sceptical about our experiments and surprised both by their students' success and by their enjoyment of the tasks. Our findings suggest that the difficulty of learning and performing arithmetic stems from the demands of constructing and operating on representations of exact number: representations beyond the limits of precision of nonsymbolic number representations[19]. When the requirement for exact precision is removed, children are able to solve symbolic addition and subtraction problems even in the absence of relevant instruction. A wide socio-economic range of children, tested in varying settings, used nonsymbolic number representations to solve symbolic problems involving quantities that fell within the scope of those preexisting representations. Arithmetic instruction may be enriched by building on this competence.

METHODS SUMMARY

In the first addition experiment, five-year-old children ($n = 20$) were given verbal problems with no visual accompaniment in which the comparison quantity differed from the sum by a 2:3 ratio. In the remaining addition experiments, children were presented with verbal problems accompanied by images of characters with sets labelled by Arabic numerals (Fig. 1a). The comparison numbers were varied to present problems at three ratios: 4:7, 4:6, and 4:5. Children in the second experiment ($n = 20$) were drawn from a highly educated community and tested in the laboratory; those in the third experiment ($n = 37$) were drawn from a lower middle-class community and tested in school. In the final experiment, children ($n = 29$) were given tests of symbolic subtraction and comparison, using verbal problems and images (Figs 1b, c). All types of problems used the same comparison quantities as the problems in past studies of nonsymbolic addition[8]. The problems were designed either to prevent or to allow tests for guessing strategies.

For each experiment, one-sample two-tailed *t*-tests compared performance to chance (50%) and one-way linear-trend analysis of variance (ANOVA) tested for the effect of ratio on performance. Performance on experiment 1 was compared to performance on a nonsymbolic addition task[8] by a one-way ('Experiment') between-groups ANOVA, with age as covariate. Performance across experiments was compared by 2 ('Experiment') by 3 ('Ratio') mixed-factor analyses of variance. Subtraction and comparison performance were compared by a 2 ('Condition') by 3 ('Ratio') repeated-measures ANOVA.

A list of problems used in the experiments, a description of the tests of school achievement and their relation to performance of symbolic approximate arithmetic, and a description of all the tests for guessing strategies are given in the Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 December 2006; accepted 18 April 2007.

1. Baroody, A. J. & Dowker, A. *The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise* (Erlbaum, Mahwah, New Jersey, 2003).

2. Geary, D. From infancy to adulthood: the development of numerical abilities. *Eur. Child Adoles. Psychiat.* **9,** 11–16 (2000).
3. Gallistel, C. R. & Gelman, R. Preverbal and verbal counting and computation. *Cognition* **44,** 43–74 (1992).
4. Dehaene, S. *The Number Sense* (Oxford Univ. Press, Oxford, UK, 1997).
5. Lemer, C., Dehaene, S., Spelke, E. & Cohen, L. Approximate quantities and exact number words: dissociable systems. *Neuropsychologia* **41,** 1942–1958 (2003).
6. Hauser, M. D. & Spelke, E. S. in *The Cognitive Neurosciences* (ed. Gazzaniga, M.) Vol. 3, 853–864 (MIT Press, Cambridge, Massachusetts, 2004).
7. McCrink, K. & Wynn, K. Large number addition and subtraction by 9-month-old infants. *Psychol. Sci.* **15,** 776–781 (2004).
8. Barth, H., La Mont, K., Lipton, J. & Spelke, E. S. Abstract number and arithmetic in preschool children. *Proc. Natl Acad. Sci. USA* **102,** 14116–14121 (2005).
9. Barth, H. *et al.* Non-symbolic arithmetic in adults and young children. *Cognition* **98,** 199–222 (2006).
10. Pica, P., Lemer, C., Izard, V. & Dehaene, S. Exact and approximate arithmetic in an Amazonian indigene group. *Science* **306,** 499–503 (2004).
11. Zur, O. & Gelman, R. Young children can add and subtract by predicting and checking. *Early Child. Res. Q.* **19,** 121–137 (2004).
12. Griffin, S. & Case, R. Evaluating the breadth and depth of training effects when central conceptual structures are taught. *Monogr. Soc. Res. Child Dev.* **61,** 83–102 (1996).
13. Lipton, J. S. & Spelke, E. S. Preschool children's mapping of number words to nonsymbolic numerosities. *Child Dev.* **76,** 978–988 (2005).
14. Le Corre, M., Van de Walle, G., Brannon, E. M. & Carey, S. Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cogn. Psychol.* **52,** 130–169 (2006).
15. Temple, E. & Posner, M. I. Brain mechanisms of quantity are similar in 5-year-old children and adults. *Proc. Natl Acad. Sci. USA* **95,** 7836–7841 (1998).
16. van Oeffelen, M. P. & Vos, P. G. A probabilistic model for the discrimination of visual number. *Percept. Psychophys.* **32,** 163–170 (1982).
17. Ashcraft, M. H. & Fierman, B. A. Mental addition in 3rd, 4th, and 6th graders. *J. Exp. Child Psychol.* **33,** 216–234 (1982).
18. LeFervre, J.-A., Smith-Chant, B. L., Hiscock, K., Daley, K. E. & Morris, J. in *The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise* (eds Baroody, A.J. & Dowker, A.) 203–228 (Erlbaum, Mahwah, New Jersey, 2003).
19. Feigenson, L., Dehaene, S. & Spelke, E. S. Core systems of number. *Trends Cogn. Sci.* **8,** 307–314 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Katz and R. Lizcano for help with data collection. This work was supported by a ROLE grant from the National Science Foundation (to E.S.S.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.K.G. (camilla.gilmore@nottingham.ac.uk) or E.S.S. (spelke@wjh.harvard.edu).

METHODS

Verbal addition experiment. Twenty children (9 male, mean age 5 years 10 months, ranging from 5 years 5 months to 6 years 6 months) were asked questions of the form ''If you had twenty-four stickers and I gave you twenty-seven more, would you have more or less than thirty-five stickers?'' Eight questions involved large numbers (in the range 5 to 98) selected so that the sum was larger on half the trials and differed from the comparison number by a ratio of 2:3. Two control questions (not analysed) involving small numbers and familiar sums ($2 + 2$ versus 20; $10 + 10$ versus 5) were included to maintain motivation and ensure children were on task. The children received no feedback but general encouragement throughout.

Computer-based symbolic addition. Twenty children (11 male, mean age 5 years 10 months, ranging from 5 years 4 months to 6 years 1 month) were given arithmetical problems presented both visually on a computer screen (see Fig. 1a) and verbally by the experimenter. Initially, two characters appeared and were named on screen. On one example trial, the experimenter stated ''Sarah has fifteen candies'' as a bag displaying the appropriate Arabic numeral appeared above the character on the left. Next, a second bag displaying an Arabic numeral appeared above the same character and the experimenter stated ''and she gets nineteen more''. Finally, a differently coloured bag displaying an Arabic numeral appeared above the other character and the experimenter stated ''John has fifty-one candies'' and asked ''Who has more?''. Twenty-four problems presented large numbers in the range 5 to 58, selected such that the sum was greater than the comparison number on half the trials and differed from it by one of three ratios—4:7, 4:6 and 4:5—on eight trials each. Three control problems (not analysed) involved small numbers and familiar sums. The session was split into three blocks and the characters and items (candies, toys, cookies) changed between blocks to maintain the children's interest in the task. Children were encouraged to guess, were given no feedback, and received stickers after each block of trials. Following the completion of the task 12 children were asked two probe exact questions. They were shown again the stimuli for two questions which they had solved correctly and were asked ''If Sarah has fifteen candies, and she gets nineteen more, how many does she have altogether?''. The sessions were videotaped to allow children's response times to be measured.

Computer-based addition in a school setting. Thirty-seven children (15 male, mean age 5 years 11 months, ranging from 5 years 5 months to 6 years 5 months) were recruited from a public kindergarten in an area of mid-to-low socioeconomic status; 37.6% of the students in the school lived in families whose income level met the state's qualifications for free or reduced-price lunch. Children were tested during school hours, in the kindergarten hallway. The procedure was otherwise the same as the computer-based addition in the laboratory.

Computer-based subtraction and comparison. Twenty-nine children (9 male, mean age 5 years 11 months, ranging from 5 years 4 months to 6 years 4 months) first completed a subtraction task and then a comparison task. The tasks were the same as in the preceding addition experiments except as follows. For the subtraction task, the experimenter said ''Sarah has fifty-five candies'' as the bag bearing an Arabic numeral appeared above the left-most character. Then a second bag then moved into view from behind the first and moved to the side as the experimenter said ''and she gives twenty-one of the candies away''. Animations were used because pilot-testing revealed that children had difficulty understanding a subtraction transformation when the operands were presented statically. The children were given four small-number practice trials (not analysed) followed by 24 experimental trials, matched to the addition trials of the computer-based addition studies so that the comparison values were the same and the difference in the computer-based subtraction was equal to the sum in the computer-based addition. Because the quantities involved in a subtraction problem are necessarily larger than those of an equivalent addition problem with the same result, a subset of the addition and subtraction problems were also matched for the numbers involved in the operation. Half the addition and subtraction problems used smaller versus larger numbers, and the small subtraction problems were matched with the large addition problems for the size of the initial quantity (for example, addition $20 + 16$; subtraction $20 - 8$). Trials were split into blocks as in the previous studies. After the last trial, 20 children were asked two exact subtraction questions of the form ''If Sarah has A candies and she gives B of them away, how many does she have left?''.

For the comparison task, a red bag appeared above the first character and the children were told ''Sarah has fifty-one candies'', then a blue jar appeared above the second character and the children were told ''Paul has thirty cookies''; finally a red bag appeared above the third character and the children were told ''John has thirty-four candies'' and were asked ''who has more candies, Sarah or John?''. An irrelevant numerical quantity was presented between the numbers to be compared in order to equalize the presentation of number words across tasks. The children were asked four small-number practice trials (not analysed) followed by 12 experimental trials in which the two values to be compared matched the sums/differences and comparison numbers for four of the eight problems at each ratio in the previous experiments. These trials were split into two blocks.

LETTERS

Universal physical responses to stretch in the living cell

Xavier Trepat[1], Linhong Deng[1,2], Steven S. An[1,3], Daniel Navajas[4], Daniel J. Tschumperlin[1], William T. Gerthoffer[5], James P. Butler[1] & Jeffrey J. Fredberg[1]

With every beat of the heart, inflation of the lung or peristalsis of the gut, cell types of diverse function are subjected to substantial stretch. Stretch is a potent stimulus for growth, differentiation, migration, remodelling and gene expression[1,2]. Here, we report that in response to transient stretch the cytoskeleton fluidizes in such a way as to define a universal response class. This finding implicates mechanisms mediated not only by specific signalling intermediates, as is usually assumed, but also by non-specific actions of a slowly evolving network of physical forces. These results support the idea that the cell interior is at once a crowded chemical space[3] and a fragile soft material in which the effects of biochemistry, molecular crowding and physical forces are complex and inseparable, yet conspire nonetheless to yield remarkably simple phenomenological laws. These laws seem to be both universal and primitive, and thus comprise a striking intersection between the worlds of cell biology and soft matter physics.

Soft materials such as tomato ketchup, shaving foam and toothpaste tend to fluidize when subjected to shear[4–7], as do granular materials including sugar in a bowl, coffee beans in a chute[8] and even certain geophysical strata during an earthquake[9]; each transforms from a solid-like to a fluid-like phase, stiffness falls, and the material flows. Underlying microscopic stress-bearing elements, or clusters of elements, interact with neighbours to form a network of force transmission, but how flow is initiated and the nature of energy barriers that must be overcome remain the subject of much current attention[5–9].

The response of a living cell to transient stretch would seem to be a different matter altogether. Very early literature shows that in response to application of a physical force the cell acutely softens (Supplementary Note 4), but more recent literature uniformly emphasizes stiffening (Supplementary Note 5)[1,10]. Nevertheless, we demonstrate here that the living cell promptly fluidizes and then slowly re-solidifies much as do the inert systems described above. Moreover, underlying structural rearrangements on the nanometre scale promptly accelerate and then slowly relax. In addition, in experiments spanning wide differences in cellular interventions, cell type and even integrative scale, these physical events conform to universal relationships.

Shear fluidization of inert matter is usually attributed to the presence of physical interactions that possess energy barriers that are so large that thermal energies by themselves are insufficient to drive microconfigurations to thermodynamic equilibrium. The material is then unable to explore its configuration space[5], and structural rearrangements become limited by long-lived microconfigurations in which the system becomes trapped. If these microconfigurations were metastable, then their longevity could depend upon agitation energy of some non-thermal origin. In the case of living cells, one such source of non-thermal agitation is ATP-dependent conformational changes of proteins[11], which release energy of about $20k_BT$ per event, where k_B is Boltzmann's constant and T is temperature, whereas another is energy injected into the system by stretch.

To test this last idea, we developed a novel experimental system in which we could subject the adherent human airway smooth muscle (HASM) cell to a transient isotropic biaxial stretch–unstretch manoeuvre of 4 seconds duration with zero residual macroscale strain. We could then monitor, on the nanometre scale, cell mechanical properties, remodelling dynamics and their changes (Methods; Supplementary Fig. 1; Supplementary Note 2).

Stiffness after stretch relative to stiffness of the same cell immediately before was denoted G'_n. When no stretch was applied, this fractional stiffness did not change, but immediately after cessation of a single transient stretch G'_n promptly decreased and then slowly recovered (Fig. 1a). These responses varied systematically with the amplitude of the imposed stretch, but little with the number of imposed stretch cycles (Supplementary Fig. 2). Immediately after stretch cessation, the phase angle $\delta = \tan^{-1}(G''/G')$ promptly increased and then slowly recovered (Fig. 1b), where for a hookean

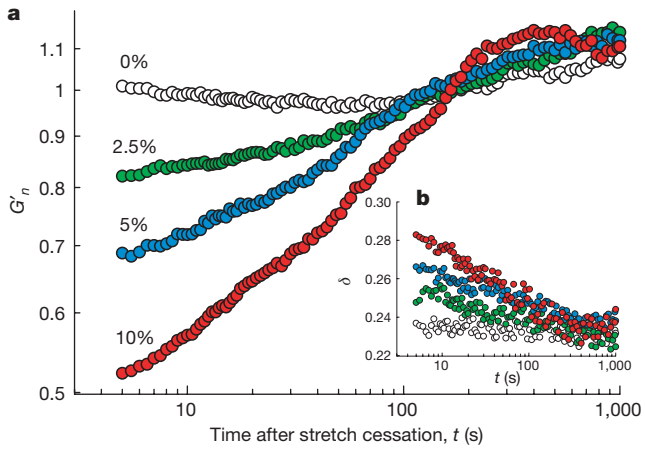


Figure 1 | A single transient stretch drives fractional stiffness G'_n down and the phase angle δ up, indicating fluidization of the cytoskeleton. **a,** Evolution of G'_n of HASM cells after a single transient stretch of 0% (no stretch, open circles), 2.5% (green), 5% (blue) and 10% (red). The response of each bead was normalized to its pre-stretch value. **b,** Evolution of the phase angle after stretch application. Compare with Box 1 in Supplementary Note 7.

[1]Program in Molecular and Integrative Physiological Sciences, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [2]'111 project' Laboratory of Biomechanics and Tissue Repair, Bioengineering College, Chongqing University, Chongqing 400044, China. [3]Division of Physiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21205, USA. [4]Unitat de Biofísica i Bioenginyeria, Universitat de Barcelona-IDIBAPS, Ciber Enfermedades Respiratorias, and Institut de Bioenginyeria de Catalunya, 08036 Barcelona, Spain. [5]Department of Pharmacology, School of Medicine, University of Nevada, Reno, Nevada 89557, USA.

solid $\delta = 0$ and for a newtonian fluid $\delta = \pi/2$. In the living cell $0.15 < \delta < 0.50$, thus placing the living cell closer to the solid-like state, and δ is virtually invariant with changes of frequency, thus setting cytoskeleton rheology within the paradigms of structural damping and scale-free dynamics[12–17]. These prompt changes establish that shear tended to fluidize the cell, and did so in a manner comparable to the effect of shear on soft materials including colloidal glasses, emulsions and pastes[4,5] (Supplementary Note 7). However, fluidization in response to transient stretch contrasts with strain-stiffening behaviour that is observed in response to sustained stretch of cells[15] or reconstituted crosslinked actin gels[18,19]; in Supplementary Note 6 we reconcile these seemingly contradictory behaviours.

To assess the robustness of these responses, we pre-treated cells with an extensive set of mechanistically distinct drugs. These interventions caused expected changes in baseline material properties (Supplementary Table 1). Despite wide differences in baseline values, each cell could serve as its own pre-stretch control. Across the panel of interventions, fluidization–resolidification responses to stretch were similar in quality but markedly disparate in magnitude and time course (Fig. 2a). When F-actin was stabilized with jasplakinolide, stretch caused the largest fractional decrease in stiffness and displayed the fastest recovery, whereas when F-actin was depolymerized with latrunculin A, stretch caused the smallest fractional decrease of stiffness and a relatively slow recovery. Inhibition of the myosin light chain kinase with ML7 blocked contractile activation as expected (Supplementary Fig. 3), but the time course of G'_n remained almost unchanged. Similarly, when extracellular calcium was chelated with EGTA to prevent calcium influx through stretch-activated channels, the time course of G'_n remained largely unchanged. In contrast, ATP depletion caused stiffness recovery to slow dramatically. In all experimental conditions, the phase angle showed a rapid increase followed by a slow decrease (Fig. 2b).

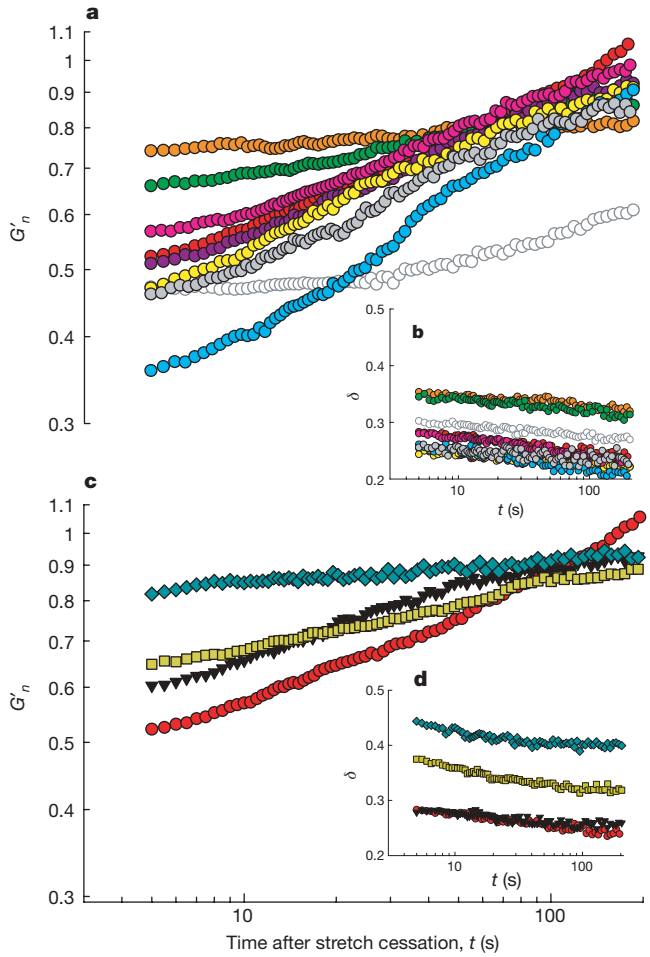


Figure 2 | A broad variety of cell systems were fluidized by a transient stretch of 10% amplitude. **a**, **b**, G'_n (**a**) and δ (**b**) of pharmacologically treated HASM cells after application of a single transient stretch of 10% amplitude (see Methods and Supplementary Table 1 for pre-stretch baseline values and treatment details). Groups are latrunculin A (orange), DBcAMP (green), ML7 (10 min incubation, bright pink; 45 min incubation, dark pink), histamine (yellow), EGTA (grey), jasplakinolide (bright blue), ATP depletion (open symbols), and untreated cells (red). G'_n (**c**) and δ (**d**) of MDCK (blue diamonds), HBE (yellow squares), HLF (black triangles) and HASM (red circles). Compare with Box 1 in Supplementary Note 7.

To assess the generality of these results, we also evaluated human lung fibroblasts (HLF), Madin–Darby canine kidney epithelial cells (MDCK) and human bronchial epithelial cells (HBE). Again, we found responses that were similar in quality but disparate in magnitude and time course (Fig. 2c, d).

A striking unification of these diverse responses was established when we focused on the prompt stiffness reduction G'_n (assessed at the earliest measurable time point, $t = 5$ s) and its initial rate of recovery α (assessed from the fit of G'_n to t^{α} for the first 30 s of response). Despite the broad diversity of drug interventions and cellular systems, when G'_n (at $t = 5$ s) was plotted against the pre-stretch value of the phase angle (δ_0), all data collapsed onto a single unifying relationship (Fig. 3a). Similarly, when the rate of stiffness recovery was plotted against δ_0 another master relationship was defined, although ATP depletion fell off that relationship (Fig. 3b). Moreover, at the level of an isolated bovine airway smooth muscle tissue strip mounted in a muscle bath, maximally activated, and then stretched using a servo-controller[20], data fell onto the very same relationships as did single cells in culture (Fig. 3).

Although these wide ranges of cellular systems, interventions and integrative scales might have led to a quagmire of inconsistencies, they instead unveiled a pattern of consistency (Fig. 3). The closer the system was to the solid-like state before being subjected to transient stretch, the greater was the extent of its fluidization and, except in the case of ATP depletion, the faster was its subsequent resolidification (Fig. 3). Behaviour of this kind is crudely reminiscent of that observed in inert soft materials such as hard-sphere colloids and is predicted by coarse-grained trap models of soft glassy rheology (Fig. 3c; Supplementary Note 7).

To test this interpretation further, we made direct observations of molecular-scale structural rearrangements. Cells were subjected to 10% transient stretch–unstretch, but instead of measuring forced bead motions using optical magnetic twisting cytometry (OMTC), as above, we measured spontaneous nanoscale bead motions and used them as a direct index of the rate of molecular-scale structural rearrangements[16]. The evolution of mean square bead displacement (MSD) on the nanometre scale—both before the transient stretch and at different waiting times (t_w) after stretch cessation—showed that when no stretch was applied the MSD evolved as we have reported previously[16,21] (Fig. 4). But when a transient stretch was applied, the rate of remodelling kinetics accelerated promptly and by more than an order of magnitude; comparatively, macromolecular mobility in unstretched cells was markedly retarded. However, as t_w increased, those kinetics progressively slowed, but relaxed more slowly than any exponential process[5–7] (Fig. 4b). This constellation of out-of-equilibrium features (Figs 1–4) represents the strongest evidence yet available to suggest slow relaxation of a glassy phase (Supplementary Note 7).

The conventional understanding of cytoskeleton dynamics has been based on the ideas that physical forces act to stiffen the network through both passive mechanical strain-stiffening[18,19] and active signalling-mediated reinforcement[1,10]. The results presented here show this viewpoint to be incomplete. Rather than merely triggering biochemical signalling cascades, as is usually assumed, cell stretch is seen to set into motion ongoing physical events in cell signalling that are not limited to the initiating upstream molecular transducers: physical forces seem to be more than a trigger. The absence of molecular

specificity in the early events of this process is highlighted by the fact that over wide ranges of systems and circumstances, the abilities of a cell to fluidize suddenly in response to stretch and to resolidify subsequently (Figs 3 and 4) seem to be insensitive to molecular details, and instead depend solely on the proximity of the cell to a solid-like state before the stretch (δ_0). This simple result is remarkable, but as more data accrue the cases in which this universality class is violated are likely to be most instructive.

These findings of universality and non-specificity, taken together with the striking analogy to dynamics in inert glassy systems (Fig. 3c; Supplementary Note 7), imply ongoing actions of a network of slowly evolving physical forces, but it remains unclear if these observations can be interpreted in terms of the onset of a non-equilibrium phase transition controlled by external stress that separates a jammed phase from a flowing phase[5,22]. Despite the underlying variety of molecular mechanisms, all such glassy systems are thought to have one feature in common—structural rearrangements that are slow, localized and inelastic—and the applicability of such a point of view is justified by the universality of the phenomenology, including inert matter[4–7,23], proteins[24], cells[16,25,26] and integrated tissues[20]. In the case of physical interactions between cytoskeletal molecules, such inelastic rearrangements might include disruption or unfolding of crosslinking proteins, resolution of steric constraints, detachment of myosin crossbridges, rupture of hydrogen bonds or cytoskeleton filaments, or actions of force-dependent capping proteins controlling filament polymerization[27].

While fluidization of living and inert systems clearly differ (Fig. 3c), the constellation of out-of-equilibrium features displayed by the cytoskeleton of the living cell is seen to be rich, nontrivial and unexplained, and would appear to describe a glassy matrix close to a glass transition. We now have firm phenomenological evidence, moreover, that dynamics in the cytoskeleton of the living adherent cell revolve around the master parameter δ_0 (or, equivalently, x; Supplementary Note 7), which sets the power-law rheology exponent[12,13], the rate of nanoscale structural rearrangements and their relaxation (Fig. 4)[16], the extent of fluidization in response to stretch (Figs 2 and 3a) and the rate of subsequent resolidification (Fig. 3b). In turn, this master parameter δ_0 is set by cytoskeletal tension (pre-stress)[28].

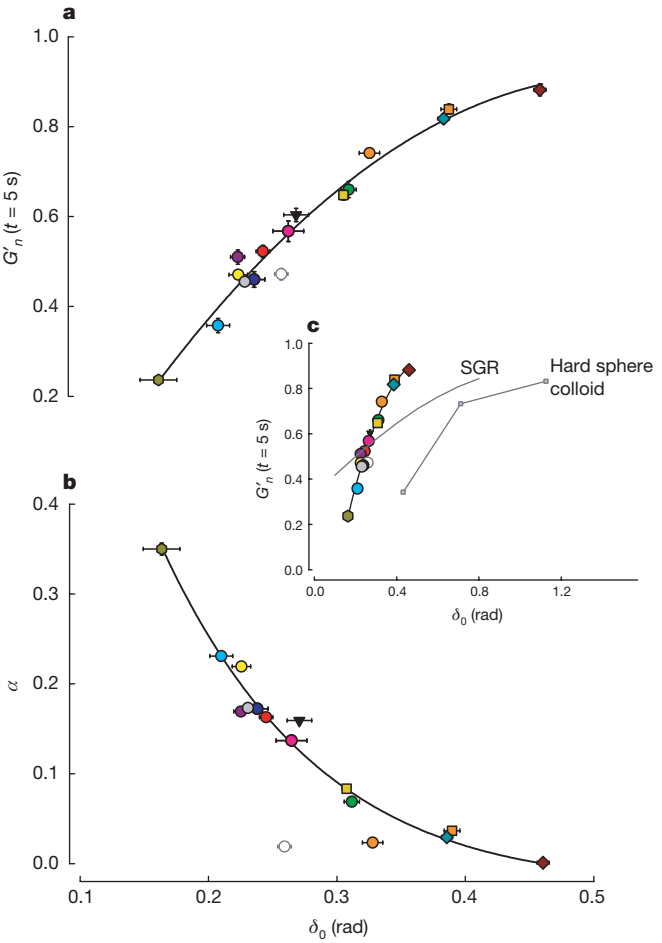


Figure 3 | Two unifying relationships describe the response to stretch of a broad variety of cell systems. In every case, the closer the system was to the solid-like state ($\delta_0 = 0$) before being subjected to transient stretch, the greater was the extent of its fluidization and, except for the case of ATP depletion, the faster was its subsequent recovery. Master curves of G'_n at the earliest time point recorded after stretch (**a**) and of the initial rate of stiffness recovery α versus the pre-stretch phase angle δ_0 (**b**). α was assessed by fitting a power-law $G'_n \propto t^{\alpha}$ to the first 30 s of response after stretch cessation. Error bars indicate standard errors. When plotted again over the full range possibilities (**c**), cells are seen to lie much closer to the solid-like ($\delta_0 = 0$) than the fluid-like ($\delta_0 = \pi/2$) state. In response to shear of similar magnitude, cells show a fluidization response comparable to but to the left of hard sphere colloids (data adapted from ref. 4). Soft glassy rheology theory[5] (Supplementary Note 7) captures these trends but substantially underestimates sensitivity to changes of δ_0. n values are given in Supplementary Table 1. Colours are as in Fig. 2 with the addition of HASM PBS (dark blue), HBE Latrunculin A (orange squares), MDCK cytochalasin D (brown diamonds) and BASM tissue (green hexagons).

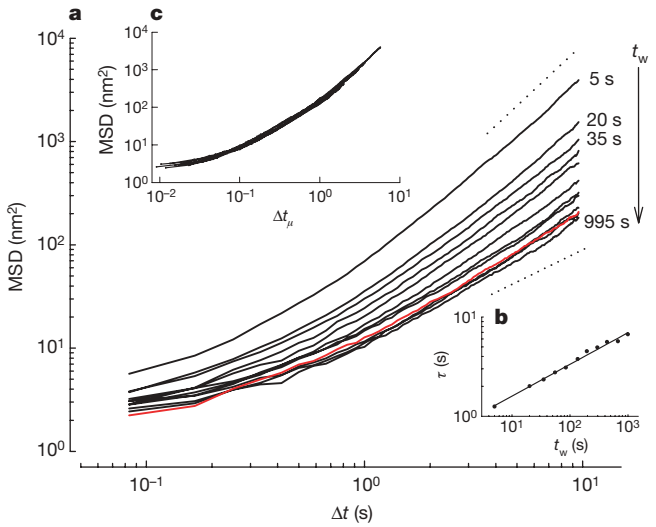


Figure 4 | Structural relaxation takes place on timescales that grow with the time elapsed since the application of stretch and is slower than any exponential process. **a**, Spontaneous motions of beads bound to HASM cells at different t_w after stretch cessation ($n = 1{,}062$ beads). Waiting times are 5, 20, 35, 55, 85, 135, 195, 295, 435, 665 and 995 s from top to bottom. The red line is the MSD before stretch application. The dashed lines indicate diffusion exponents of 1 and 2. **b**, To characterize the progressive slowing of rearrangement kinetics, we defined a time τ at which $\mathrm{MSD}(\tau) = d^2$, where d was taken as an arbitrary threshold and τ thus represented the average time required for a bead to move (diffuse) a distance d. For any value of d, we found that τ increased with t_w as a power law $\tau \propto t_w^{\mu}$ with $\mu \approx 0.3$, indicating that the decay was slower than any exponential process, and that within the experimental time window no steady state was achieved. Data are shown for $d^2 = 100\ \mathrm{nm}^2$ and the solid line is a fit to a power law with exponent $\mu = 0.32$. **c**, After rescaling the time axis using $\Delta t_\mu = \Delta t / t_w^{\mu}$ with $\mu = 0.32$, all data collapsed onto a master curve. This indicates that the kinetics at each waiting time were self-similar. In inert soft glassy materials, such slowing of rearrangement kinetics as well as the absence of a steady state is referred to as physical ageing and μ is identified as the ageing coefficient. Physical ageing can be interrupted by injection of mechanical energy through shear; shear drives inelastic structural rearrangements[7,23], in which case it is presumed that elements can then 'hop' out of the deep energy wells in which they are trapped, erase system memory, and push the system farther from thermodynamic equilibrium. In inert soft materials these events reset system evolution to some earlier time and for that reason are called physical rejuvenation[6,7].

As such, two major concepts in cytoskeletal biology that are each understood in its own right to be highly unifying are now seen to be linked intimately to one another—cytoskeletal tension on the one hand and glassy dynamics on the other. Although a mechanism explaining why this might be so is unknown[29] (Supplementary Notes 7, 8), these phenomena, taken together, define the most primitive features of the cytoskeletal phenotype, namely, its abilities to deform, to contract and to remodel, and might have arisen as early as the cytoskeleton itself, about two billion years ago[30] (Supplementary Note 9).

We traditionally think about molecular interactions within the cytoskeleton as being highly specific, whereas in a crowded[3] glassy phase the nature of molecular interactions and their rate of progression become highly constrained, severely regularized, and non-specific (Figs 3 and 4; Supplementary Note 8). Conventional descriptions of physical interactions based upon viscoelasticity, the fluctuation-dissipation theorem, specific signalling cascades, dilute solution chemistry, and even emerging notions of protein interaction maps, fail to account for these dynamics. Because physical interactions are now seen to play out within a glassy phase, they will have to be rethought within a rather different conceptual perspective.

METHODS SUMMARY

Cell culture. HASM, HLF, HBE and MDCK cells were grown on collagen I-coated silastic substrates (Flexcell) for 5–14 days, depending on the cell type. Cells were allowed to reach confluence before being serum-deprived for experiments.

Measurement of cell rheology. We measured cell rheology using OMTC[12,13] (Supplementary Notes 2). Ferrimagnetic beads (4.5 μm) coated with a synthetic Arg–Gly–Asp (RGD)-containing peptide are allowed to bind integrins on the cell surface and become tightly anchored to the cytoskeleton through focal contacts. The beads are permanently magnetized in the horizontal plane of the cell culture and subsequently twisted in an oscillatory magnetic field with frequency 0.75 Hz. The twisting field causes each bead to rotate towards alignment with the oscillatory field and, as result, a weak mechanical torque is applied to the cell. The complex modulus of the cells (G^*) is computed from the Fourier transform of the applied mechanical torque (T) and of the resulting lateral bead displacement (D):

$$G^* = G' + jG'' = \tilde{T}/\tilde{D} \quad (1)$$

where * denotes a complex number, the tilde overbar denotes the Fourier domain, and $j^2 = -1$. For each experimental condition, data are reported as medians of the bead populations (215–719 beads on a similar number of cells per experimental condition).

Spontaneous bead motions. Spontaneous bead motions were measured by tracking the position of the centroid of each bead. Data were recorded during time intervals of 10 s starting at different waiting times t_w after stretch cessation. We computed the MSD of each bead i as:

$$\mathrm{MSD}_i(\Delta t, t_w) = \langle (x_i(t+\Delta t, t_w) - x_i(t, t_w))^2 \rangle \quad (2)$$

where Δt is the time lag, x is the bead coordinate, and brackets indicate an average over t. The distribution of the $\mathrm{MSD}(\Delta t, t_w)$ from bead to bead was approximately log-normal. Accordingly, data are reported as the median of the bead population.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 November 2006; accepted 10 April 2007.

1. Vogel, V. & Sheetz, M. Local force and geometry sensing regulate cell functions. *Nature Rev. Mol. Cell Biol.* **7,** 265–275 (2006).
2. Ingber, D. E. & Tensegrity, I. I. How structural networks influence cellular information processing networks. *J. Cell Sci.* **116,** 1397–1408 (2003).
3. Minton, A. P. How can biochemical reactions within cells differ from those in test tubes? *J. Cell Sci.* **119,** 2863–2869 (2006).
4. Mason, T. G. & Weitz, D. A. Linear viscoelasticity of colloidal hard sphere suspensions near the glass transition. *Phys. Rev. Lett.* **75,** 2770–2773 (1995).
5. Sollich, P., Lequeux, F., Hebraud, P. & Cates, M. E. Rheology of soft glassy materials. *Phys. Rev. Lett.* **78,** 2020–2023 (1997).
6. Cloitre, M., Borrega, R. & Leibler, L. Rheological aging and rejuvenation in microgel pastes. *Phys. Rev. Lett.* **85,** 4819–4822 (2000).
7. Viasnoff, V. & Lequeux, F. Rejuvenation and overaging in a colloidal glass under shear. *Phys. Rev. Lett.* **89,** 065701 (2002).
8. Corwin, E. I., Jaeger, H. M. & Nagel, S. R. Structural signature of jamming in granular media. *Nature* **435,** 1075–1078 (2005).
9. Johnson, P. A. & Jia, X. Nonlinear dynamics, granular media and dynamic earthquake triggering. *Nature* **437,** 871–874 (2005).
10. Matthews, B. D., Overby, D. R., Mannix, R. & Ingber, D. E. Cellular adaptation to mechanical stress: role of integrins, Rho, cytoskeletal tension and mechanosensitive ion channels. *J. Cell Sci.* **119,** 508–518 (2006).
11. Lau, A. W., Hoffman, B. D., Davies, A., Crocker, J. C. & Lubensky, T. C. Microrheology, stress fluctuations, and active behavior of living cells. *Phys. Rev. Lett.* **91,** 198101 (2003).
12. Fabry, B. *et al.* Scaling the microrheology of living cells. *Phys. Rev. Lett.* **87,** 148102 (2001).
13. Fabry, B. *et al.* Time scale and other invariants of integrative mechanical behavior in living cells. *Phys. Rev. E* **68,** 041914 (2003).
14. Alcaraz, J. *et al.* Microrheology of human lung epithelial cells measured by atomic force microscopy. *Biophys. J.* **84,** 2071–2079 (2003).
15. Trepat, X. *et al.* Viscoelasticity of human alveolar epithelial cells subjected to stretch. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **287,** L1025–L1034 (2004).
16. Bursac, P. *et al.* Cytoskeletal remodelling and slow dynamics in the living cell. *Nature Mater.* **4,** 557–561 (2005).
17. Deng, L. *et al.* Fast and slow dynamics of the cytoskeleton. *Nature Mater.* **5,** 636–640 (2006).
18. Gardel, M. L. *et al.* Elastic behavior of cross-linked and bundled actin networks. *Science* **304,** 1301–1305 (2004).
19. Storm, C., Pastore, J. J., MacKintosh, F. C., Lubensky, T. C. & Janmey, P. A. Nonlinear elasticity in biological gels. *Nature* **435,** 191–194 (2005).
20. Fredberg, J. J. *et al.* Airway smooth muscle, tidal stretches, and dynamically determined contractile states. *Am. J. Respir. Crit. Care Med.* **156,** 1752–1759 (1997).
21. Hoffman, B. D., Massiera, G., Van Citters, K. M. & Crocker, J. C. The consensus mechanics of cultured mammalian cells. *Proc. Natl Acad. Sci. USA* **103,** 10259–10264 (2006).
22. Miguel, M. C. & Zapperi, S. Materials science. Fluctuations in plasticity at the microscale. *Science* **312,** 1151–1152 (2006).
23. Bulatov, V. V. & Argon, A. S. A stochastic-model for continuum elastoplastic behavior. 2. A study of the glass-transition and structural relaxation. *Model. Simul. Mater. Sci. Eng.* **2,** 185–202 (1994).
24. Brujic, J., Hermans, R. I., Walther, K. A. & Fernandez, J. M. Single-molecule force spectroscopy reveals signatures of glassy dynamics in the energy landscape of ubiquitin. *Nature Phys.* **2,** 282–286 (2006).
25. Moazzam, F., DeLano, F. A., Zweifach, B. W. & Schmid-Schonbein, G. W. The leukocyte response to fluid stress. *Proc. Natl Acad. Sci. USA* **94,** 5338–5343 (1997).
26. Yap, B. & Kamm, R. D. Mechanical deformation of neutrophils into narrow channels induces pseudopod projection and changes in biomechanical properties. *J. Appl. Physiol.* **98,** 1930–1939 (2005).
27. Kozlov, M. M. & Bershadsky, A. D. Processive capping by formin suggests a force-driven mechanism of actin polymerization. *J. Cell Biol.* **167,** 1011–1017 (2004).
28. Stamenovic, D., Suki, B., Fabry, B., Wang, N. & Fredberg, J. J. Rheology of airway smooth muscle cells is associated with cytoskeletal contractile stress. *J. Appl. Physiol.* **96,** 1600–1605 (2004).
29. Rosenblatt, N., Alencar, A. M., Majumdar, A., Suki, B. & Stamenovic, D. Dynamics of prestressed semiflexible polymer chains as a model of cell rheology. *Phys. Rev. Lett.* **97,** 168101 (2006).
30. Kirschner, M. W. & Gerhart, J. C. *The Plausibility of Life: Resolving Darwin's Dilemma* (Yale Univ., New Haven, 2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements These studies were supported by grants from National Institutes of Health and from the Spanish Ministries of Education and Science and Health. We thank R. Panettieri for providing cells, and R. Farré, D. Fletcher, F. Ritort and V. Viasnoff for discussions.

Author Contributions X.T. and J.J.F. designed research and wrote the manuscript. J.P.B. conducted the theoretical analysis. X.T. and D.N. designed and implemented the experimental system. X.T., L.D. and S.S.A. optimized experimental conditions and treatments. W.T.G. and D.J.T. helped to design experimental protocols and interpret data. D.J.T. provided cells and reagents. X.T. performed all stretch experiments and data analysis. J.J.F. oversaw the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.J.F. (jeffrey_fredberg@harvard.edu).

METHODS

Cell culture. HASM cells were provided by R. Panettieri (Univ. Pennsylvania). Cells were grown on collagen I-coated silastic substrates in Ham's Nutrient Mixture F-12 medium supplemented with 10% fetal bovine serum (FBS), 100 U ml^{-1} penicillin, 100 $\mu g\,ml^{-1}$ streptomycin, 200 $\mu g\,ml^{-1}$ amphotericin B, 12 mM NaOH, 1.6 mM $CaCl_2$, 2 mM L-glutamine and 25 mM HEPES. After reaching confluence (7–14 days after plating), cells were serum deprived and supplemented with 5.7 $\mu g\,ml^{-1}$ insulin and 5 $\mu g\,ml^{-1}$ human transferrin. These conditions have been shown to maximize the expression of smooth-muscle-specific proteins and restore the contractile phenotype of smooth muscle cells. Experiments were performed 48 h after serum deprivation. HBE cells (16-HBE-14o- line) were cultured in minimum essential medium (MEM) with Earle's salts supplemented with 10% FBS, 2 mM L-glutamine, 100 U ml^{-1} penicillin, and 100 $\mu g\,ml^{-1}$ streptomycin. MDCK cells (strain II) were cultured in MEM with Earle's salts supplemented with 5% FBS, 2 mM L-glutamine, 100 U ml^{-1} penicillin and 100 $\mu g\,ml^{-1}$ streptomycin. Adult HLF (LL-24 line) were cultured in F12K medium supplemented with 15% FBS, 100 U ml^{-1} penicillin and 100 g ml^{-1} streptomycin. The cells were serum deprived on the day of experiments. All experiments were conducted in confluent cell monolayers.

Pharmacological interventions. The following pharmacological interventions were used to modulate the cytoskeleton; histamine (HASM cell contraction, 50 μM), (dibutyryl cAMP, HASM cell relaxation, 1 mM), ML7 (inhibition of the myosin light chain kinase, 30 μM), jasplakinolide (F-actin stabilization, 0.1 μM), latrunculin-A (disruption of F-actin through sequestration of actin monomers, 0.1 μM) and cytochalasin D (disruption of F-actin through filament capping, 0.1 μM). Extracellular calcium was chelated by incubating cells in calcium-free physiological saline solution with 1 mM EGTA. ATP was depleted by incubating cells in phosphate-buffered saline (PBS) with NaN_3 (2 mM) and deoxyglucose (10 mM) for 45–60 min. ATP concentrations after depletion were 10% of those in control samples.

Microbead and cell preparation. Ferrimagnetic microbeads (4.5 μm in diameter), produced in our laboratory, were coated with a peptide containing the sequence Arg–Gly–Asp (RGD). The ligands were adsorbed onto the bead surface (50 μg ligand per mg beads in 1 ml of carbonate buffer, pH 9.4) by overnight incubation at 4 °C.

The coated beads were added for 20 min at 37 °C to allow binding to the receptors on the cell surface. The RGD-coated beads attach to HASM cells through integrin receptors (mainly β_1 integrins) and become strongly anchored to the cytoskeleton (Supplementary Notes 2). The wells were washed twice with serum-free medium to remove any unbound beads. 20–40 min later, the well was mounted on the set-up described below for experiments.

Cell stretching system. Experiments were performed with a custom-built stretching device that produced uniform biaxial deformations of the flexible substrate upon which cells were firmly adherent (Supplementary Fig. 1a). The substrate was placed on a lubricated O-ring positioned over the objective of an inverted optical microscope. Applying a transient vacuum pressure to the underside of the outer region of the substrate stretches the membrane over the O-ring, and consequently, stretches the attached cells. The rising and falling times of the transient stretch are set by a resistor–capacitor pneumatic filter inserted between the sample and a vacuum source. To measure cell mechanics immediately before and after stretch application, two pairs of coaxial coils were coupled to the stretching device, one to magnetize the beads and another to twist them.

Vol 447|31 May 2007|doi:10.1038/nature05851

LETTERS

Rewiring cellular morphology pathways with synthetic guanine nucleotide exchange factors

Brian J. Yeh[1,2]*, Robert J. Rutigliano[3]*, Anrica Deb[2], Dafna Bar-Sagi[3,4] & Wendell A. Lim[2]

Eukaryotic cells mobilize the actin cytoskeleton to generate a remarkable diversity of morphological behaviours, including motility, phagocytosis and cytokinesis. Much of this diversity is mediated by guanine nucleotide exchange factors (GEFs) that activate Rho family GTPases—the master regulators of the actin cytoskeleton[1–3]. There are over 80 Rho GEFs in the human genome (compared to only 22 genes for the Rho GTPases themselves), and the evolution of new and diverse GEFs is thought to provide a mechanism for linking the core cytoskeletal machinery to a wide range of new control inputs. Here we test this hypothesis and ask if we can systematically reprogramme cellular morphology by engineering synthetic GEF proteins. We focused on Dbl family Rho GEFs, which have a highly modular structure common to many signalling proteins[4,5]: they contain a catalytic Dbl homology (DH) domain linked to diverse regulatory domains, many of which autoinhibit GEF activity[2,3]. Here we show that by recombining catalytic GEF domains with new regulatory modules, we can generate synthetic GEFs that are activated by non-native inputs. We have used these synthetic GEFs to reprogramme cellular behaviour in diverse ways. The GEFs can be used to link specific cytoskeletal responses to normally unrelated upstream signalling pathways. In addition, multiple synthetic GEFs can be linked as components in series to form an artificial cascade with improved signal processing behaviour. These results show the high degree of evolutionary plasticity of this important family of modular signalling proteins, and indicate that it may be possible to use synthetic biology approaches to manipulate the complex spatio-temporal control of cell morphology.

Rho family GTPases are central signalling molecules in the regulation of the actin cytoskeleton[1] (Fig. 1a). These proteins are conformational switches that exist in GDP- and GTP-bound states; however, only the GTP-bound state actively transduces signal to downstream effectors. Cycling between states is primarily controlled by opposing enzymes: GTPase activating proteins promote hydrolysis of bound GTP to GDP (inactivation), whereas GEFs promote exchange of bound GDP for GTP (activation). The three canonical members of the Rho family—Cdc42, Rac1 and RhoA—stimulate the distinct morphological outputs of protrusive filopodia (thin actin microspikes), protrusive lamellipodia (broad membrane ruffles) and contractile actin:myosin filaments, respectively.

As an initial target for rewiring GTPase signalling, we attempted to reprogramme Dbl family GEFs (Fig. 1b) so that their activity was controlled by protein kinase A (PKA), a well-characterized prototypical kinase[6] (Fig. 2a). We first designed a PKA-sensitive autoinhibitory module, inspired by natural examples[7], that consisted of a PDZ (PSD95, Dlg, ZO-1) domain–peptide interaction pair that could be disrupted by PKA phosphorylation. The syntrophin PDZ domain recognizes short carboxy-terminal peptide motifs (consensus sequence (R/K)E(S/T)xψ-COOH; ψ denotes aliphatic residues)[8], which are close in sequence to the ideal PKA substrate (RRRRSIIFI)[9]. A hybrid sequence (RRRESIV-COOH) could serve both as an interaction ligand for the syntrophin PDZ domain and

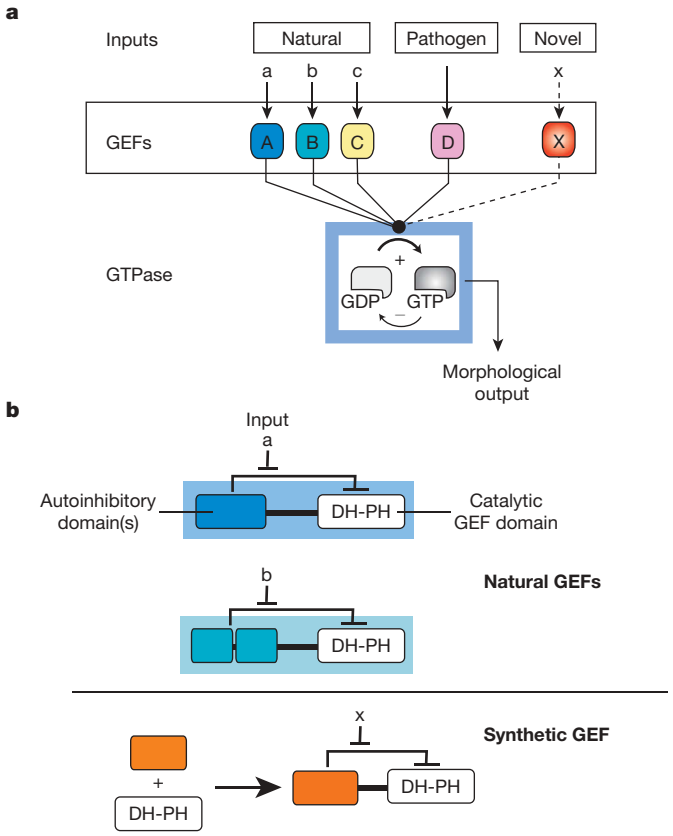


Figure 1 | GEFs link diverse inputs to Rho GTPase modules that control cell morphology. a, GEFs functionally connect signalling inputs to activation of Rho GTPases, which regulate morphology of the actin cytoskeleton. Some bacterial pathogens encode GEFs that activate host GTPases[30]. Synthetic GEFs could, in principle, mediate new connections in living cells. **b,** The largest family of Rho GEFs are Dbl-related proteins, which share a catalytic DH-PH core. In many cases, adjacent modular domains mediate autoinhibitory interactions that can be disrupted by specific inputs. Here we exploit this modular structure to construct synthetic GEFs.

[1]Chemistry and Chemical Biology Graduate Program, University of California, San Francisco, San Francisco, California 94158-2517, USA. [2]Department of Cellular and Molecular Pharmacology and the Cell Propulsion Lab, UCSF/UCB NIH Nanomedicine Development Center, University of California, San Francisco, San Francisco, California 94158-2517, USA. [3]Department of Molecular Genetics and Microbiology, School of Medicine, State University of New York at Stony Brook, Stony Brook, New York 11794, USA. [4]Department of Biochemistry, New York University School of Medicine, New York, New York 10016, USA.
*These authors contributed equally to this work.

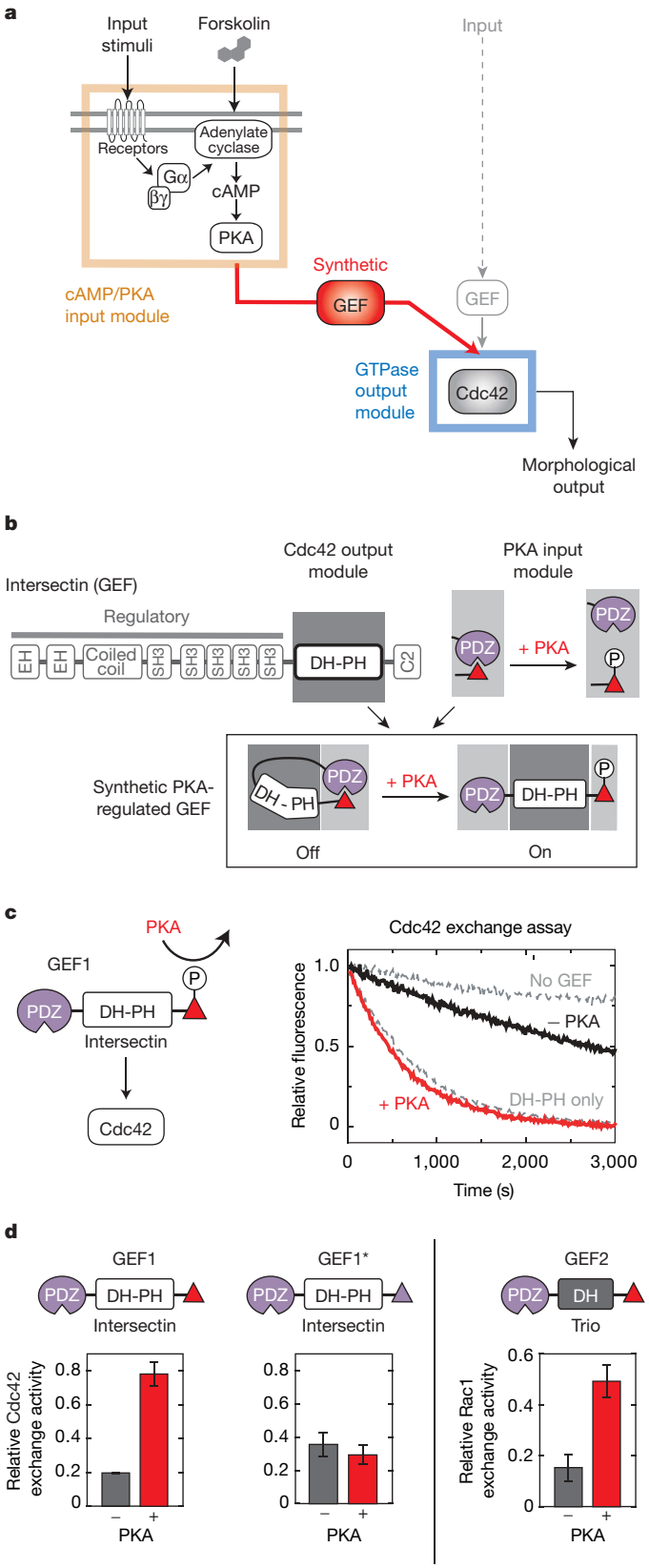


Figure 2 | Modular recombination yields PKA-responsive synthetic GEFs. **a**, We attempted to engineer GEFs that link PKA signalling to specific cytoskeletal changes. **b**, PKA-sensitive GEFs were constructed by fusing Dbl homology (DH) GEF output modules with a PKA input module composed of the syntrophin PDZ domain and a peptide that binds the PDZ domain and is a PKA substrate. **c**, *In vitro* assay of GEF1 showing activation by PKA. Dissociation of fluorescent mant-GDP from Cdc42 was measured in the presence of no GEF or constitutively active intersectin DH-PH (dotted lines), GEF1 (solid black line), or GEF1 pre-treated with PKA (red line). **d**, Activities of synthetic GEFs (relative to intersectin DH-PH or Trio DH). GEF1 and GEF2 were basally repressed, but were activated by PKA. GEF1* contains a mutation that abolishes phosphorylation by PKA, but retains binding to the syntrophin PDZ. Error bars represent s.d. of three experiments. Substrate specificities of GEF1 and GEF2 were identical to those of their respective parental DH proteins, intersectin and Trio (Supplementary Table 4).

a PKA substrate. Most importantly, we found that phosphorylation by PKA disrupted binding to the PDZ domain (Supplementary Fig. 1).

To build a Cdc42 GEF that could be activated by PKA, we fused this PKA-sensitive PDZ–peptide interaction module to the Dbl homology–pleckstrin homology (DH-PH) catalytic core from intersectin (Itsn1)—a Cdc42-specific Dbl family member, the catalytic activity of which is normally regulated by autoinhibitory SH3 domains[10,11] (Fig. 2b). We refer to this construct as GEF1 (see Supplementary Tables 1 and 2 for details of all synthetic GEFs). In an *in vitro* Cdc42 nucleotide exchange assay, GEF1 was repressed relative to the constitutively active DH-PH fragment (<20% activity), indicating that the intramolecular PDZ interaction sterically occluded or conformationally disrupted the DH-PH domain (Fig. 2c, d and Supplementary Fig. 2). Phosphorylation of GEF1 by PKA relieved repression, increasing Cdc42 exchange activity (Fig. 2c, d and Supplementary Fig. 2). For a control, we mutated the peptide to a sequence that could still bind the PDZ domain but could not be phosphorylated by PKA. A construct bearing this mutation (GEF1*) was still repressed, but was not activated by PKA (Fig. 2d and Supplementary Fig. 2).

To test if the PKA regulatory module could be transferred to another GEF, we replaced the intersectin DH-PH with the amino-terminal DH domain of Trio, which preferentially activates Rac1 (GEF2)[12]. GEF2 was also repressed *in vitro* (relative to the Trio DH domain alone), and could be activated by PKA (Fig. 2d and Supplementary Fig. 2). A control construct bearing a non-phosphorylatable peptide (GEF2*) could not be activated by PKA (Supplementary Fig. 3).

In total, we fused the PDZ-peptide module to the DH and/or DH-PH fragments of five Dbl family members with varying GTPase specificities (including intersectin and Trio), and tested their activity *in vitro*. All seven constructs tested showed some degree of repression under basal conditions, and four out of seven were activated by PKA (Supplementary Table 3). No attempts were made to optimize autoinhibitory affinity, domain orientation or interdomain linker lengths, and it is likely that such efforts would improve activation of the three remaining synthetic GEFs[13].

To test if these synthetic GEF proteins could create new functional signalling linkages *in vivo*, we introduced GEF1 and GEF2 into cells by microinjection. We first tested the effect of microinjecting the unregulated catalytic GEF modules into the REF52 fibroblast cell line (Fig. 3a). As expected, microinjection of the Trio DH domain led to a constitutive Rac1-associated lamellipodial phenotype. Microinjection of the intersectin DH-PH module yielded a constitutive Cdc42-associated filopodial phenotype in a large fraction of cells; however, a significant but inconsistent fraction of these cells showed an alternative rounded phenotype that is distinct from filopodia and lamellipodia (Supplementary Fig. 4). Co-injection of additional Cdc42 with the intersectin DH-PH resulted only in cells with filopodia, alleviating this dual phenotype problem. Thus, to simplify phenotypic scoring, we co-injected the relevant GTPases in all of the following experiments. We estimate that we are increasing the cellular concentration of the specific GTPases by approximately two-fold, which has no morphological effect in the absence of the GEF domain. This method has previously been used to clarify scoring of GEF-induced phenotypes[14].

Injection of GEF1 into REF52 cells resulted in a new PKA-activated filopodial response. After microinjecting the purified proteins, we tested the cellular response to stimulation with increasing doses of forskolin, a pharmacological activator of PKA[15] (Fig. 3b). When GEF1 was injected into cells, even in the absence of PKA stimulation, there was a weak background activity; 14% of the cells showed filopodia, probably owing to somewhat leaky repression of GEF activity. However, this phenotype was much weaker than that observed with injection of an equivalent amount of the unregulated DH-PH module (>95% of cells with filopodia). Most importantly, filopodia were stimulated in a dose-dependent fashion as a function of forskolin concentration such that >60% of the cells had filopodia at the highest forskolin concentrations tested (Fig. 3c). Furthermore, induction of filopodia was observed within minutes of forskolin addition to cells pre-injected with GEF1 (Supplementary Movie), demonstrating the rapid timescale of response with protein-based networks that do not require transcription and translation. Forskolin treatment of cells lacking GEF1 (injected only with Cdc42) led to a small background stimulation of filopodia (~20%), indicating that there is only a weak endogenous linkage between PKA and filopodia formation in REF52 cells. As an important control, we observed no significant stimulation of filopodia in cells microinjected with GEF1*, which is autoinhibited but cannot be activated by PKA. These results imply that the strong stimulation of filopodia is the result of a new, functional signalling connection mediated directly by the engineered GEF1 protein.

Similarly, injection of GEF2 into REF52 cells resulted in a PKA-inducible lamellipodial response. Injection of GEF2 and Rac1 had little basal effect on the cells (4% of cells with lamellipodia); however, treatment with forskolin resulted in a dose-dependent increase in the number of cells with lamellipodia (to >60%) (Fig. 3d). Activation of lamellipodia also occurred within minutes of stimulation (data not shown). Cells injected with GEF2* showed no significant lamellipodial response to forskolin (Supplementary Fig. 3). Thus, both synthetic GEFs are capable of mediating linkages between the endogenous PKA signalling pathway and Rho GTPase-mediated morphological rearrangements in live cells.

Many complex behaviours observed in living cells are mediated by multiple signalling proteins that do not function alone, but instead are linked into more complex multistep pathways[16]. For example, the canonical GTPase Ras can activate multiple effectors, including the Rac1 GEF, Tiam1 (ref. 17). Thus, we asked whether we could link synthetic GEFs with specifically engineered input–output linkages into a two GTPase cascade in which PKA would activate Cdc42, and Cdc42 would in turn activate Rac1 (Fig. 4a). GEF1 could provide the connection between PKA and Cdc42; however, the second step required a Cdc42-responsive autoinhibitory module, which we extracted from the signalling protein N-WASP as a GTPase-binding domain (GBD) that recognizes a short central domain (C). The GBD–C interaction is normally involved in autoinhibition of N-WASP, and can be disrupted by activated Cdc42 (refs 18–21). We fused the GBD-C module to the Trio DH domain, producing a Rac1-specific GEF that is activated by Cdc42 (GEF3). *In vitro* analysis of GEF3 showed that its Rac1 exchange activity was regulated by Cdc42, as expected (data not shown), providing further evidence for the flexibility of this overall framework for engineering diverse signalling linkages.

Co-injection of GEF1 and GEF3 (along with Cdc42 and Rac1) into REF52 cells resulted in a new signalling cascade: PKA stimulation by forskolin ultimately led to Rac1 activation and a lamellipodial phenotype (Fig. 4b). Almost no filopodial response was observed, perhaps because lamellipodia tend to be dominant over filopodia, and because much of the activated Cdc42 may be sequestered by binding to GEF3 instead of other effectors. To confirm that signal is passing through both synthetic GEFs, we disrupted each individual component. GEF3 was selectively disrupted by a small deletion in the GBD that blocks binding to Cdc42(GTP) but does not affect autoinhibition (GEF3*)[20,21]. When GEF1 and GEF3* were injected into REF52 cells, forskolin treatment led to only the activation of Cdc42, resulting in robust formation of filopodia (Fig. 4b). Similarly, GEF1* is a

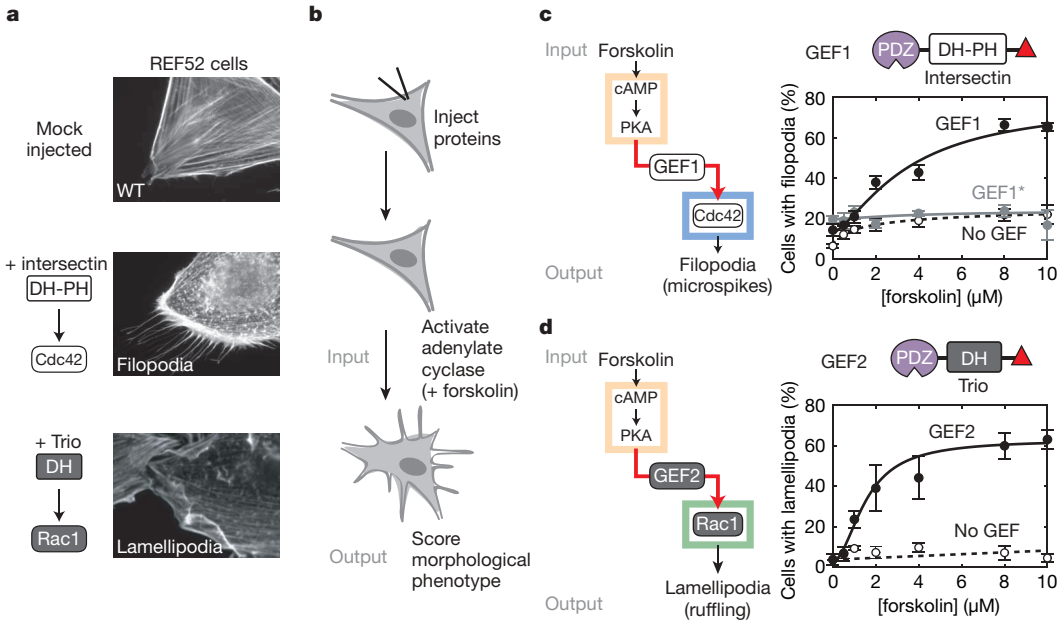


Figure 3 | Synthetic GEFs generate new PKA-dependent morphological changes in cells. **a**, Microinjection of constitutively active intersectin DH-PH (and Cdc42) induced filopodia in REF52 cells. Constitutively active Trio DH (and Rac1) induced lamellipodia. **b**, After microinjecting synthetic GEFs, cells were treated with forskolin, which activates endogenous PKA. Morphological response was scored by counting cells exhibiting filopodia or lamellipodia. **c**, Filopodia were stimulated by forskolin treatment of cells injected with GEF1 (solid black line). Forskolin treatment of cells lacking GEF1 led only to a weak background stimulation of filopodia (dashed line). Filopodia were not significantly stimulated in cells injected with GEF1* (grey line), which is autoinhibited but cannot be activated by PKA. Data points represent mean ± s.d. of three experiments (>50 cells scored per experiment), and were fit to a conventional Hill equation. **d**, Injection with GEF2 allowed stimulation of lamellipodia by forskolin (solid line). Little or no response was observed in cells lacking GEF2 (dashed line). Error bars as in **c**.

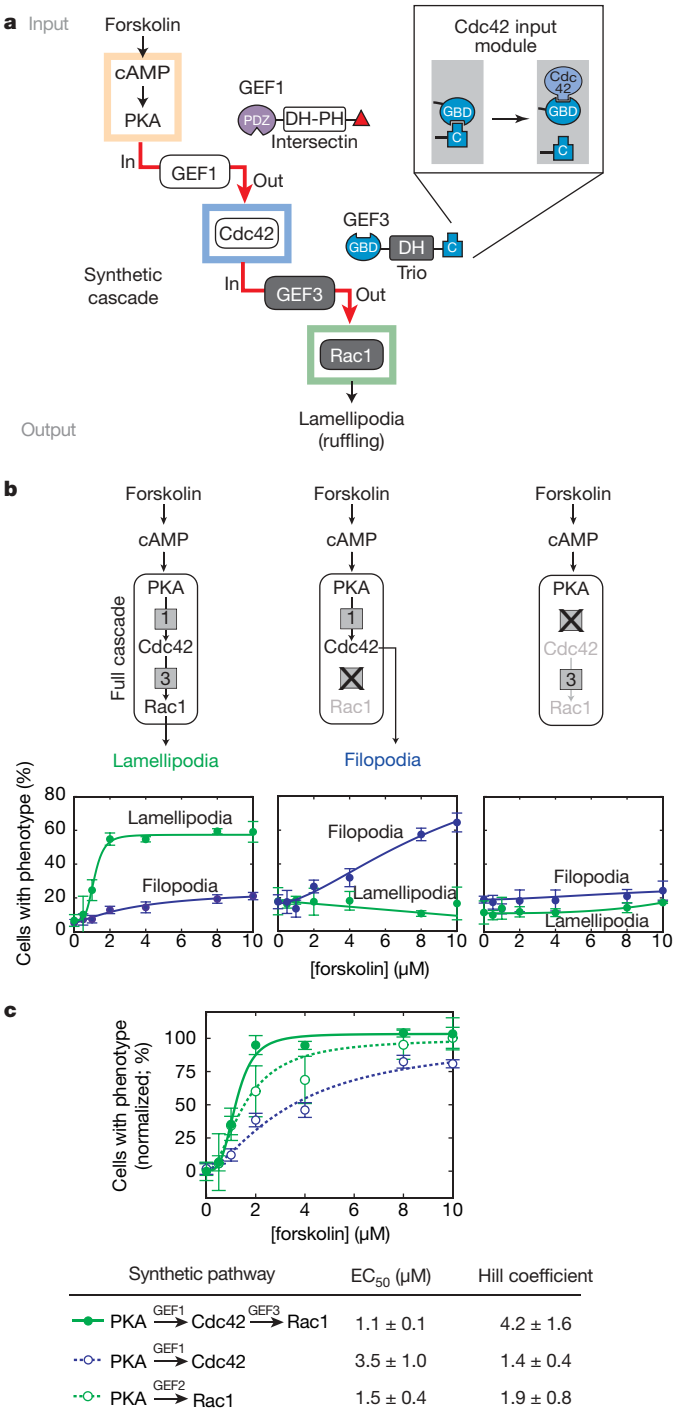


Figure 4 | Two synthetic GEFs can be linked in series to form a higher order cascade. a, GEF1 and GEF3 form a cascade in which PKA activates Cdc42, which in turn activates Rac1. GEF3 is composed of the Trio DH domain and a Cdc42 input module extracted from N-WASP. **b**, Co-injection of GEF1 and GEF3 resulted in a functioning cascade (left panel). Forskolin induced lamellipodia (green curve) with very little induction of filopodia (blue curve). Forskolin treatment of cells co-injected with GEF1 and GEF3* (which cannot respond to Cdc42) resulted only in filopodia (middle panel). Cells co-injected with GEF1* and GEF3 showed no significant filopodial or lamellipodial response (right panel). Data points represent mean ± s.d. of three experiments (>50 cells scored per experiment), and were fit to a conventional Hill equation. **c**, Comparison of the GEF1–GEF3 cascade (solid green line) to direct single-GEF circuits mediated by GEF1 (dashed blue line) or GEF2 (dashed green line). Data were normalized to lower and upper baselines obtained from fits to the Hill equation. EC_{50}, effector concentration for half-maximum response. Error bars as in **b**.

variant of GEF1 that cannot be phosphorylated and activated by PKA. As predicted, cells injected with GEF1* and GEF3 showed no activation of either Cdc42 or Rac1 (no significant filopodial or lamellipodial response) on forskolin treatment. Together, these results imply that GEF1 and GEF3 form a functional signalling cascade that links PKA to a lamellipodial response.

The GEF1–GEF3 cascade demonstrated several properties that distinguished it from the direct, single GEF circuits (Fig. 4c and Supplementary Fig. 5). First, the synthetic cascade had dampened noise: there was a ~2-fold reduction in basal response (no forskolin stimulation), both in terms of filopodial and lamellipodial output, when GEF3 was introduced downstream of GEF1. Second, the cascade seemed to amplify response within a certain range of stimulation. In the direct PKA→Cdc42 circuit (GEF1 only), the amount of Cdc42 activated by 1–2 μM forskolin was insufficient to mount a significant filopodial response. However, in the GEF1–GEF3 cascade, this low Cdc42 activation was sufficient to activate GEF3, producing a stronger Rac1-mediated lamellipodial response. Third, the cascade is ultrasensitive—it had a sharp activation threshold with an apparent Hill coefficient (n_H) of >4, despite the fact that its individual components respond in a linear (michaelian) fashion. The increased ultrasensitivity of the cascade is consistent with theoretical and experimental studies that compared pathways with increasing number of steps[16,22]. Although individual signalling proteins can exhibit non-linear behaviours[13], these simple synthetic GEFs can be linked into higher order architectures that begin to show complex emergent properties.

Here we demonstrate that Rho GEFs provide a flexible framework for engineering novel signalling pathways. Modular recombination allows the expansion of GTPase control relationships beyond those generated through evolution. GTPases regulate many biological processes (nuclear trafficking, endocytosis, and so on)[23]; thus, such approaches could be applied to manipulate these processes. These findings, along with related studies, demonstrate that modular protein signalling components can be engineered in a relatively facile manner, indicating that it may be possible to apply synthetic biology approaches to generate cells with precisely engineered target behaviours[5,24–27]. Although there has been significant progress in engineering transcriptional networks in living cells[28,29], there are comparatively fewer examples of synthetic signal transduction networks. These protein-based networks are important because they mediate many of the rapid and spatially precise responses in cells, including complex properties such as cell shape and movement. The ability to manipulate these properties will be critical for engineering cells with diverse therapeutic and biotechnological applications.

METHODS

For detailed information on all methods see Supplementary Information.

Synthetic GEFs. Proteins (sequence details in Supplementary Tables 1, 2 and 3) were expressed as hexahistidine fusions in *Escherichia coli* BL21(DE3)RIL, and purified by chromatography on Ni-NTA resin (Qiagen).

***In vitro* nucleotide exchange assays.** Dissociation of *N*-methylanthraniloyl (mant)-GDP from Cdc42 (qualitative assays) or association of mant-GDP with GTPases (quantitative assays) were measured using a SpectraMax Gemini XS (Molecular Devices) fluorescence multi-well plate reader (25 °C; excitation, 360 nm; emission, 440 nm).

Microinjection experiments. Rat embryo fibroblasts (REF52) were grown as sub-confluent monolayers overnight and serum-starved for 24 h before injection. Proteins were injected into the cytosol of cells using an Eppendorf 5246 pressure system and an Eppendorf 5171 microinjector. GEFs were injected at the molar equivalent of 2 mg ml^{-1} intersectin DH-PH, and associated GTPases were co-injected at 0.5 mg ml^{-1} (concentrations are in the needle). Injected cells were incubated for 30 min at 37 °C, treated with the indicated concentration of forskolin for 30 min, and allowed to recover from any deleterious effects of the drug for 30 min. Cells were then fixed, stained with rhodamine-phalloidin, and mounted onto glass slides. Morphological phenotypes were scored in a blind fashion (without knowledge of the experimental condition). Cells showing at least 5 protrusive spikes were scored positive for filopodia, and cells that had dense peripheral actin staining were scored positive for lamellipodia (no cells

were observed with both filopodia and lamellipodia). The percentage of cells with each phenotype was calculated by dividing the number of cells with the scored phenotype by the total number of cells scored.

Received 19 December 2006; accepted 17 April 2007.
Published online 21 May 2007.

1. Jaffe, A. B. & Hall, A. Rho GTPases: biochemistry and biology. *Annu. Rev. Cell Dev. Biol.* **21,** 247–269 (2005).
2. Hoffman, G. R. & Cerione, R. A. Signaling to the Rho GTPases: networking with the DH domain. *FEBS Lett.* **513,** 85–91 (2002).
3. Rossman, K. L., Der, C. J. & Sondek, J. GEF means go: turning on RHO GTPases with guanine nucleotide-exchange factors. *Nature Rev. Mol. Cell Biol.* **6,** 167–180 (2005).
4. Pawson, T. & Nash, P. Assembly of cell regulatory systems through protein interaction domains. *Science* **300,** 445–452 (2003).
5. Bhattacharyya, R. P., Remenyi, A., Yeh, B. J. & Lim, W. A. Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu. Rev. Biochem.* **75,** 655–680 (2006).
6. Walsh, D. A. & Van Patten, S. M. Multiple pathway signal transduction by the cAMP-dependent protein kinase. *FASEB J.* **8,** 1227–1236 (1994).
7. Kim, E. & Sheng, M. PDZ domain proteins of synapses. *Nature Rev. Neurosci.* **5,** 771–781 (2004).
8. Wiedemann, U. *et al.* Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides. *J. Mol. Biol.* **343,** 703–718 (2004).
9. Songyang, Z. *et al.* Use of an oriented peptide library to determine the optimal substrates of protein kinases. *Curr. Biol.* **4,** 973–982 (1994).
10. Snyder, J. T. *et al.* Structural basis for the selective activation of Rho GTPases by Dbl exchange factors. *Nature Struct. Biol.* **9,** 468–475 (2002).
11. Zamanian, J. L. & Kelly, R. B. Intersectin 1L guanine nucleotide exchange activity is regulated by adjacent src homology 3 domains that are also involved in endocytosis. *Mol. Biol. Cell* **14,** 1624–1637 (2003).
12. Debant, A. *et al.* The multidomain protein Trio binds the LAR transmembrane tyrosine phosphatase, contains a protein kinase domain, and has separate rac-specific and rho-specific guanine nucleotide exchange factor domains. *Proc. Natl Acad. Sci. USA* **93,** 5466–5471 (1996).
13. Dueber, J. E., Mirsky, E. A. & Lim, W. A. Engineering synthetic signaling proteins with ultrasensitive input/output contol. *Nature Biotech.* (in the press) (2007).
14. Nimnual, A. S., Yatsula, B. A. & Bar-Sagi, D. Coupling of Ras and Rac guanosine triphosphatases through the Ras exchanger Sos. *Science* **279,** 560–563 (1998).
15. Seamon, K. B. & Daly, J. W. Forskolin: a unique diterpene activator of cyclic AMP-generating systems. *J. Cyclic Nucleotide Res.* **7,** 201–224 (1981).
16. Ferrell, J. E. Jr. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends Biochem. Sci.* **21,** 460–466 (1996).
17. Yamauchi, J., Miyamoto, Y., Tanoue, A., Shooter, E. M. & Chan, J. R. Ras activation of a Rac1 exchange factor, Tiam1, mediates neurotrophin-3-induced Schwann cell migration. *Proc. Natl Acad. Sci. USA* **102,** 14889–14894 (2005).
18. Miki, H., Sasaki, T., Takai, Y. & Takenawa, T. Induction of filopodium formation by a WASP-related actin-depolymerizing protein N-WASP. *Nature* **391,** 93–96 (1998).
19. Rohatgi, R. *et al.* The interaction between N-WASP and the Arp2/3 complex links Cdc42-dependent signals to actin assembly. *Cell* **97,** 221–231 (1999).
20. Kim, A. S., Kakalis, L. T., Abdul-Manan, N., Liu, G. A. & Rosen, M. K. Autoinhibition and activation mechanisms of the Wiskott–Aldrich syndrome protein. *Nature* **404,** 151–158 (2000).
21. Prehoda, K. E., Scott, J. A., Mullins, R. D. & Lim, W. A. Integration of multiple signals through cooperative regulation of the N-WASP–Arp2/3 complex. *Science* **290,** 801–806 (2000).
22. Hooshangi, S., Thiberge, S. & Weiss, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl Acad. Sci. USA* **102,** 3581–3586 (2005).
23. Colicelli, J. Human RAS superfamily proteins and related GTPases. *Sci. STKE* **2004,** RE13 (2004).
24. Dueber, J. E., Yeh, B. J., Chak, K. & Lim, W. A. Reprogramming control of an allosteric signaling switch through modular recombination. *Science* **301,** 1904–1908 (2003).
25. Howard, P. L., Chia, M. C., Del Rizzo, S., Liu, F. F. & Pawson, T. Redirecting tyrosine kinase signaling to an apoptotic caspase pathway through chimeric adaptor proteins. *Proc. Natl Acad. Sci. USA* **100,** 11267–11272 (2003).
26. Park, S. H., Zarrinpar, A. & Lim, W. A. Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. *Science* **299,** 1061–1064 (2003).
27. Inoue, T., Heo, W. D., Grimley, J. S., Wandless, T. J. & Meyer, T. An inducible translocation strategy to rapidly activate and inhibit small GTPase signaling pathways. *Nature Methods* **2,** 415–418 (2005).
28. Sprinzak, D. & Elowitz, M. B. Reconstruction of genetic circuits. *Nature* **438,** 443–448 (2005).
29. Voigt, C. A. Genetic parts to program bacteria. *Curr. Opin. Biotechnol.* **17,** 548–557 (2006).
30. Patel, J. C. & Galan, J. E. Manipulation of the host actin cytoskeleton by Salmonella—all in the name of entry. *Curr. Opin. Microbiol.* **8,** 10–15 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. C. Anderson, A. Arkin, H. Bourne, J. Dueber, G. Kapp, J. Kardon, M. Nyako, and members of the Lim and Bar-Sagi laboratories for assistance and discussion. This work was supported by grants from the NIH (D.B.-S. and W.A.L.), the Packard Foundation (W.A.L.), and the Rogers Family Foundation (W.A.L.). B.J.Y. was supported by a Post-Graduate Scholarship from NSERC and R.J.R. was supported by an NIH-NCI Cancer Biochemistry and Cell Biology training grant.

Author Contributions B.J.Y., R.J.R., D.B.-S. and W.A.L. conceived the experiments. B.J.Y. and A.D. designed and purified the constructs and performed the *in vitro* experiments. R.J.R. performed the *in vivo* experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to W.A.L. (lim@cmp.ucsf.edu) or D.B.-S. (dafna.bar-sagi@med.nyu.edu).

LETTERS

The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation

Mamta Tahiliani[1,2]*, Pinchao Mei[1]*, Rui Fang[1], Thiago Leonor[1], Michael Rutenberg[1,3], Fumiko Shimizu[1], Jing Li[1], Anjana Rao[2] & Yujiang Shi[1]

Gene transcription is critically influenced by chromatin structure and the modification status of histone tails[1]. Methylation of lysine residues in histone tails is dynamically regulated by the opposing activities of histone methyltransferases and histone demethylases[2]. Here we show that JARID1C/SMCX, a JmjC-domain-containing protein implicated in X-linked mental retardation and epilepsy[3,4], possesses H3K4 tri-demethylase activity and functions as a transcriptional repressor. An SMCX complex isolated from HeLa cells contains additional chromatin modifiers (the histone deacetylases HDAC1 and HDAC2, and the histone H3K9 methyltransferase G9a) and the transcriptional repressor REST[5], suggesting a direct role for SMCX in chromatin dynamics and REST-mediated repression. Chromatin immunoprecipitation reveals that SMCX and REST co-occupy the neuron-restrictive silencing elements in the promoters of a subset of REST target genes. RNA-interference-mediated depletion of SMCX derepresses several of these targets and simultaneously increases H3K4 trimethylation at the sodium channel type 2A (*SCN2A*) and synapsin I (*SYN1*) promoters. We propose that loss of SMCX activity impairs REST-mediated neuronal gene regulation, thereby contributing to SMCX-associated X-linked mental retardation.

We isolated an SMCX-containing protein complex from nuclear extracts of HeLa cells stably expressing doubly tagged (with Flag and haemagglutinin, HA) human SMCX (Fig. 1a). Among the polypeptides specifically associated with SMCX were: the transcriptional repressor E2F6 and its partners DP1, Max and Mga[6]; the nuclear receptor co-repressor 1 (NCOR1)[7]; the RE-1 silencing transcription factor (REST), which represses neuronal genes in non-neuronal tissues[5]; the chromatin-modifying enzymes HDAC1 and HDAC2 and G9a[8,9]; the euchromatin-associated protein HP1γ (ref. 6); the E3 ubiquitin ligases RING1 and RING2 (ref. 10); and the polycomb group protein, h-l(3)mbt-like[6]. The collective presence of these proteins points to a role for the SMCX complex in transcriptional repression.

Glycerol gradient sedimentation indicated that SMCX is a component of at least two distinct large protein complexes, containing E2F6 and REST respectively (peaks at fractions 13 and 17 are indicated by black lines above fraction numbers in Fig. 1b; uncomplexed REST and SMCX migrate at the top of the gradient in fractions 4–10). We focus here on characterization of the SMCX–REST complex. REST co-immunoprecipitated with its known partners G9a and HDAC2 (refs 8, 9), as well as Flag:HA-SMCX, RING2 and NCOR1 (Fig. 1c). Reciprocally, glutathione *S* transferase (GST)–SMCX pulled down endogenous REST from HeLa cell nuclear extracts (Supplementary Fig. 1). REST did not significantly associate with E2F6, confirming that the E2F6–SMCX and SMCX–REST complexes are distinct (Fig. 1c).

SMCX contains a JmjC domain, demonstrated to possess histone demethylase activity in other proteins[11–13]. Indeed, the SMCX complex displayed robust demethylase activity against both H3K4me3 and H3K4me2, as well as histone deacetylase activity consistent with the presence of HDAC1 and HDAC2 (Fig. 1d, panels 1, 2 and 6). Immunoblotting with a histone H3-specific antibody established that the observed decrease in methylation was not due to degradation

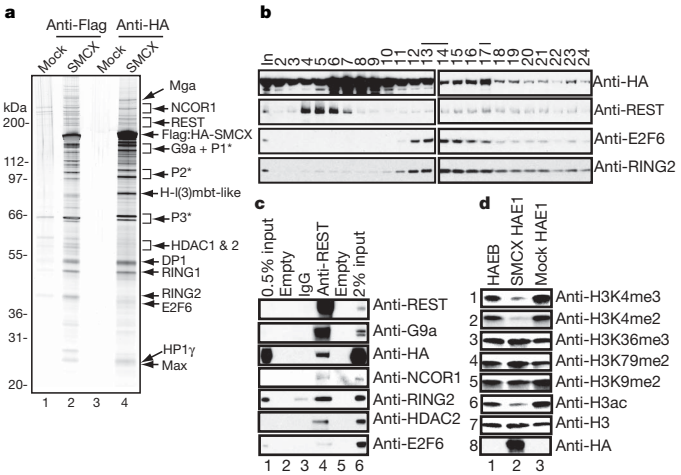


Figure 1 | The SMCX complex contains an H3K4 tri-demethylase activity. **a**, Purification and mass spectrometric analysis of polypeptides associated with SMCX. Nuclear extracts from HeLa cells stably expressing human Flag:HA tagged-SMCX (lanes 2 and 4) and untransduced HeLa cell nuclear extracts (lanes 1 and 3) were sequentially purified with Flag and HA antibody affinity resins. Associated polypeptides were detected by silver staining. P1*, P2* and P3* denote potential SMCX degradation products identified by mass spectrometry and immunoblotting. **b**, Flag-affinity-purified Flag:HA-SMCX material (SMCX FlagE1) was separated on a 10–40% glycerol gradient and then input (In) and fractions were analysed by immunoblotting. **c**, Reciprocal immunoprecipitation. FlagE1 (lane 2, Fig. 1a) was incubated with either control immunoglobulin γ, IgG (lane 3), or anti-REST antibody conjugated beads (lane 4). Input (lanes 1 and 6) and immunoprecipitates were analysed by immunoblotting. **d**, Tandem-affinity-purified SMCX material (SMCX HAE1) was tested for enzymatic activity using core histones as substrates (lane 2). Core histones were incubated with HA peptide elution buffer (HAEB) and mock purified material (Mock HAE1) for comparison (lanes 1 and 3). Histone methylation and acetylation status were determined using immunoblotting. Data are representative of at least two experiments.

[1]Division of Endocrinology, Diabetes, and Hypertension, Department of Medicine and BCMP, Brigham and Women's Hospital and Harvard Medical School, 221 Longwood Avenue, [2]Harvard Medical School and the CBR Institute for Biomedical Research, 200 Longwood Avenue, Boston, Massachusetts 02115, USA. [3]Department of Pathology, University of Florida College of Medicine, Gainesville, Florida 32610, USA.
*These authors contributed equally to this work.

(Fig. 1d, panel 7). Other modifications associated with active transcription (H3K36me3, H3K79me2) or repression (H3K9me2) were not diminished by incubation with the SMCX complex (Fig. 1d, panels 3–5).

H3K4 di- and tri-demethylase activities are intrinsic properties of SMCX (Fig. 2). Recombinant GST–SMCX (Supplementary Fig. 2a) demethylated H3K4me2 and H3K4me3 (Fig. 2a, panels 1 and 7) while other modifications (H3K4me1, H3K9me2 and H3K9me3, H3K36me2 and H3K36me3, H3K79me2) were unaffected. As a control, JMJD2C demethylated H3K9me3 and H3K36me3 (Fig. 2a, panels 3 and 9) but not H3K4me2 or H3K4me3 (Fig. 2a, panels 1 and 7)[12]. Matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) (Fig. 2b) confirmed that the chemical reaction catalysed by JmjC-containing histone demethylases results in a net loss of 14 daltons for each methyl group removed, because the loss of a methyl group is accompanied by the addition of a proton[11,13]. GST–SMCX demethylated synthetic H3 peptides bearing K4me3 and K4me2 modifications to yield K4me1 (Fig. 2b, panels 8 and 5), but did not affect H3 peptides bearing K4me1 or K9me3 modifications (Fig. 2b, panel 2, and Supplementary Fig. 2b).

Fe(II) and α-ketoglutarate are essential co-factors for JmjC-containing histone demethylases[11]. A catalytic triad (H, D/E, H) is predicted to be involved in Fe(II) coordination and is highly conserved in most JmjC-containing proteins[13,14]. A mutant GST–SMCX, in which one of the predicted iron-coordinating residues, H514, is substituted with tyrosine, was completely devoid of demethylase activity, establishing that the JmjC domain is responsible for the demethylase activity of SMCX (Fig. 2b, panels 3, 6 and 9).

Unlike LSD1, which demethylates nucleosomes as a protein complex but not a recombinant protein[15], recombinant GST–SMCX and purified SMCX complexes were unable to demethylate nucleosomes, the physiological substrates for histone modification (Supplementary Fig. 3). Nevertheless, green fluorescent protein (GFP)–SMCX expressed in human osteoblastic U2OS cells almost completely erased nucleosomal H3K4me3 marks (Fig. 3a, panel 5), whereas the catalytically inactive GFP–SMCX-H514Y was inactive (Fig. 3a, panel 8). In contrast to the robust H3K4 di-demethylase activity observed *in vitro*, the decrease in H3K4me2 staining was only partial in cells (Fig. 3a, panel 2), possibly because H3K4me2 levels reach an equilibrium determined by relative rates of production (through H3K4me3 demethylation) and consumption (conversion to H3K4me1). H3K4me1 staining was unaffected as expected (Supplementary Fig. 2c).

The SMCX complex contains HDAC 1 and 2 (Fig. 1). However, the demethylase activity of SMCX did not depend on its HDAC activity (Supplementary Fig. 3). GST–SMCX and the SMCX complex equivalently demethylated purified histones, derived from untreated HeLa cells or cells treated with the HDAC inhibitor trichostatin A (TSA) to induce histone hyperacetylation (Supplementary Fig. 3). Moreover, TSA treatment had little influence on H3K4me2/H3K4me3 demethylation in GFP–SMCX-expressing U2OS cells (Supplementary Fig. 4), nor did TSA affect the demethylase activity of the SMCX complex *in vitro* (data not shown).

Deletion of the first 69 amino acids of SMCX, which contain the JmjN domain, completely abolished the demethylase activity of GFP–SMCX in cells (Fig. 3b). This deletion (SMCX-C1), and even more extensive deletions of the amino-terminal 389 and 759 amino acids (SMCX-C2 and SMCX-C3), did not affect the predominant nuclear localization of GFP–SMCX; conversely, mutants lacking

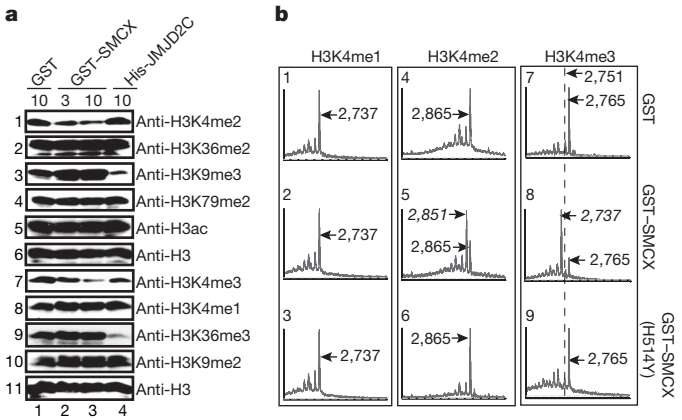


Figure 2 | Recombinant SMCX purified from insect Sf9 cells demethylates H3K4me2 and H3K4me3 *in vitro*. **a**, GST–SMCX demethylates histone H3K4me2/H3K4me3 *in vitro*. 10 μg of bulk histones were incubated with 3 or 10 μg of GST–SMCX in histone demethylase reaction buffer. Histones were incubated under identical conditions with purified GST and His-JMJD2C as controls. 20% (panels 1–6) and 80% (panels 7–11) of the reactions were run on separate gels and histone methylation status was determined by immunoblotting; the membrane with the higher amount of histones was used for less sensitive antibodies. **a**, In panel 5, GST–SMCX is not able to reduce levels of acetyl H3. **b**, MALDI-TOF MS analysis of H3K4me1/H3K4me2/H3K4me3 peptides incubated with GST, GST–SMCX or GST–SMCX(H514Y). The enzyme-to-substrate ratio is 1:15. Numbers represent the masses (in daltons) of substrate (upright) and product (italic) peptides respectively. **b**, The dotted line in panels 7–9 denotes the expected position of a peak corresponding to the production of a dimethylated peptide. Data are representative of at least three experiments.

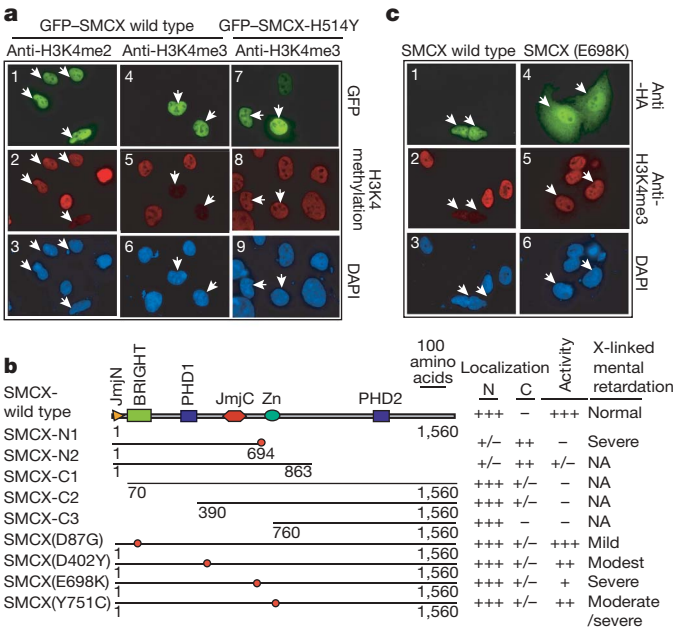


Figure 3 | SMCX demethylates di- and trimethyl H3K4 in cells. **a**, U2OS cells overexpressing GFP-SMCX show (in panels 1–3) weakened H3K4me2 staining and (in panels 4–6) a strong decrease in H3K4me3 levels. In panels 7–9, overexpression of enzymatically inactive mutant GFP–SMCX(H514Y) has no effect on H3K4me3 levels. **b**, Schematic representation of wild-type and mutant SMCX proteins along with subcellular localization and effect on H3K4me3 levels. +, enzymatically active; –, enzymatically inactive; +/–, enzymatic activity varied from cell to cell. N, nuclear subcellular localization; C, cytoplasmic subcellular localization; +/–, subcellular localization varied from cell to cell. **c**, Disease-associated mutation SMCX(E698K) affects H3K4 tri-demethylase activity *in vivo*. Panels 1–3 show that cells transfected with Flag:HA-SMCX show a strong decrease in H3K4me3 signal. Panels 4–6 show that Flag:HA-SMCX(E698K) shows nuclear localization with some cytoplasmic accumulation, and retains only minimal H3K4 tri-demethylase activity. Arrows point to transfected cells. Exposure times were held constant to allow direct comparison of staining levels. Data are representative of at least three experiments. DAPI, 4,6-diamidino-2-2 phenylindole dihydrochloride. NA, not applicable. BRIGHT, B cell regulator of IgH transcription; PHD1 and PHD2, PHD zinc finger 1 and 2; Zn, C5HC2 zinc finger.

the carboxy-terminal portion of SMCX (SMCX-N1 and SMCX-N2) were poorly localized to the nucleus. Together, these results indicate that both the JmjC domain and the JmjN domain are essential for enzymatic activity, while the C-terminal region is responsible for targeting SMCX to the nucleus.

Within the past two years, over twenty independent mutations in the *SMCX* gene have been reported in patients with mild to severe X-linked mental retardation [3,4]. None of these mutations are located in the JmjC domain, and the majority of the mutations do not occur in predicted functional domains. We selected four representative point mutations in SMCX, located within or near domains we identified as critical for the catalytic activity of SMCX, for analysis of demethylase activity *in vivo* (Fig. 3b, red dots). Wild-type and mutant SMCX proteins were transiently expressed in U2OS cells, and their effect on H3K4 tri-methylation was assessed by immunocytochemistry.

The cellular demethylase activities of this series of SMCX mutants correlated roughly with the severity of associated X-linked mental retardation (Fig. 3b). SMCX (D87G), associated with mild X-linked mental retardation, showed strong nuclear localization and minimal defects in demethylase activity *in vivo*. SMCX (D402Y) and SMCX (Y751C), associated with moderate X-linked mental retardation, were predominantly but not completely nuclear, and showed substantial but incomplete impairment of H3K4 tri-demethylase activity. SMCX (E698K), associated with severe X-linked mental retardation, yielded the most striking phenotype, causing almost no decrease in H3K4me3 staining in cells (Fig. 3b, c). This was not due to gross impairment of catalytic activity, as shown by testing the demethylase activity of GST–SMCX(E698K) *in vitro* (Supplementary Fig. 5), nor did it reflect impaired chromatin association of the E698K mutant in stably transfected HeLa cells (data not shown).

SMCX functioned as a transcriptional repressor in transient reporter assays (Supplementary Fig. 6). A fusion protein of SMCX with the GAL4 DNA-binding domain (Supplementary Fig. 6a) repressed, in a dose-dependent manner, the activity of a luciferase reporter plasmid containing tandem GAL4 binding sites and the minimal thymidine kinase promoter. The extent of repression was similar to that observed with a GAL4–LSD1 fusion protein (Supplementary Fig. 6b). The N-terminal deletion mutants (G4-SMCXΔ70 and Δ390), and the point mutant associated with severe X-linked mental retardation, G4-SMCX(E698K), were partly compromised in their ability to repress transcription compared to wild-type SMCX, as was the catalytically inactive H514Y mutant when tested at low doses (Supplementary Fig. 6d, e, f). We speculate that histone demethylase activity may not be fully responsible for repression by SMCX. HDACs do not seem to contribute significantly to SMCX-mediated repressive activity, as shown in luciferase reporter assays conducted in the presence of TSA (Supplementary Fig. 6i, j). Thus the bulk of SMCX-mediated repression may reflect the presence of other transcriptional repressors in the SMCX complex.

Because mutations in SMCX are implicated in neuronal disorders such as mental retardation and epilepsy [3,4], we focused on the functional interaction between SMCX and REST. REST binds neuron-restrictive silencing elements or RE-1 elements found in the promoters of many neuronal genes including *SYN1*, *SCN2A*, superior cervical ganglion-10 (*SCG10*), and promoter II of brain-derived neurotropic factor (*BDNF*) [5,8,16,17]. Chromatin immunoprecipitation (ChIP) assays showed that endogenous REST and tagged SMCX were both present at these neuron-restrictive silencing elements or RE-1 sites in HeLa cells stably expressing Flag:HA-SMCX, but not at the promoter of *CDC25A*, which is not a REST target gene [18], or at the coding regions of *SCN2A*, *SCG10* or a control gene RNA polymerase II (*POLII*) (Fig. 4a, b). Moreover, sequential ChIP assays showed that SMCX and REST simultaneously occupy the neuron-restrictive silencing elements of several REST target genes (*SCN2A*, *SCG10*, *BDNF* and *SYN1* promoters) but not the *CDC25A* promoter (Fig. 4c). RNA interference (RNAi)-mediated depletion of SMCX to 27% of control levels resulted in 1.8- and 2.0-fold increases in expression of *SYN1* and *SCN2A* respectively, while *β-actin* and *GAPDH* levels remained unaffected (Fig. 4d).

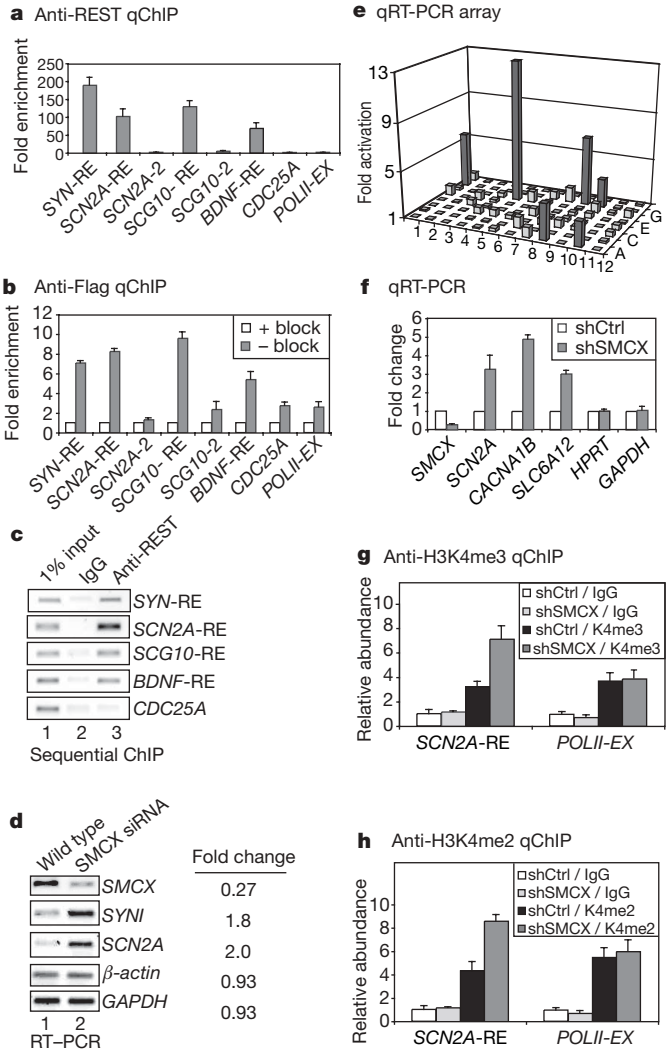


Figure 4 | SMCX regulates promoter H3K4me2/H3K4me3 levels and expression of several REST target genes. **a**, **b**, Enrichment of REST and SMCX at neuron-restrictive silencing elements or RE-1 sites (-RE) analysed by quantitative ChIP (qChIP). –EX, exon. **a**, Anti-REST qChIP, with data expressed as fold enrichment relative to IgG control antibody. **b**, Anti-Flag qChIP, with data expressed as fold enrichment relative to immunoprecipitation performed in the presence of blocking Flag peptide. **a**, **b**, Error bars represent the s.e.m. of duplicates. ±block, in the presence/absence of blocking Flag peptide. **c**, Sequential ChIP shows that SMCX and REST co-occupy the neuron-restrictive silencing elements or RE-1 sites of several REST target genes. Anti-IgG was used as a negative control for the second immunoprecipitation. *SYN1*-RE, *SCN2A*-RE, *SCG10*-RE and *BDNF*-RE are located in the promoters of these genes. *SCN2A*-2 and *SCG10*-2 are located in the introns of *SCN2A* and *SCG10*. *POLII*-EX is located in exon 24 of *POLII* and *CDC25A* is located in the promoter of *CDC25A*. **d**, Duplex RNAi-mediated depletion of SMCX derepresses several REST targets. siRNA, small interfering RNA. **e**–**h**, Experiments were performed on stable SMCX knockdown HeLa cell lines generated by transduction with SMCX shRNA (shSMCX) or control shRNA (shCtrl). **e**, Identification of SMCX target genes. Shaded genes show over 2-fold derepression. The columns of the PCR array are numbered 1–12 and alternate rows are labelled A, C, E, G. **f**, Quantitative RT–PCR confirming upregulation of targets identified in **e**. Depletion of SMCX resulted in increased H3K4me3 (**g**) and H3K4me2 (**h**) at *SCN2A*-RE and not *POLII-EX*. **f**–**h**, Data are expressed as relative abundance and is normalized to input. Error bars represent s.d. of triplicates. Data shown are representative of at least two experiments. GAPDH, glyceraldehyde-3-phosphate dehydrogenase.

We used a commercial polymerase chain reaction with reverse transcription (RT–PCR) array to examine the effect of SMCX knockdown on the expression of 84 neuronal genes (ion channels and transporters) in HeLa cells stably expressing SMCX-directed or control short hairpin (sh)RNA (Supplementary Fig. 7a). The two cell lines expressed equivalent levels of the SMCX-associated proteins G9a, REST, NCOR1 and HDAC1 (Supplementary Fig. 7b). Consistent with a role for SMCX as a transcriptional repressor, knockdown of SMCX resulted in derepression (>2-fold) of six genes (Fig. 4e, Supplementary Table 1). *SCN2A* was among these genes, confirming our previous result (Fig. 4d). The other derepressed genes fall into three categories: *CACNA1B* and *CACNA1H*, encoding calcium channels; *SCL6A3* and *SLC18A1*, encoding monoamine transporters; and *SLC6A12*, encoding a GABA transporter. For one gene selected from each group, upregulation was confirmed by RT–PCR (Fig. 4f). All six SMCX target genes have been implicated in neurological disorders such as epilepsy, autism or schizophrenia[19–25]; moreover, *CACNA1B* and *CACNA1H* are reported REST target genes[26,27].

Consistent with its demethylase activity, RNAi-mediated depletion of SMCX increased H3K4 trimethylation levels at the *SCN2A* and *SYN1* promoters without affecting global levels of H3K4me2/H3K4me3 or H3 acetylation (Fig. 4g, h, Supplementary Figs 7c, 8b). As expected, H3K4me3 levels were depleted at REST-silenced gene promoters compared to the promoter of an actively transcribed gene such as *CDC25A* (Supplementary Fig. 8a). Depletion of SMCX resulted in doubling of H3K4me2/H3K4me3 levels at the neuron-restrictive silencing element or RE-1 site of *SCN2A* (*SCN2A* RE) (Fig. 4g, h) and H3K4me3 levels at the *SYN1* RE (Supplementary Fig. 8b), with almost no effect on these modifications at *POLII*, a non-REST target gene chosen because it has H3K4me3 levels similar to that of *SCN2A*.

A central mechanism for eukaryotic gene regulation involves dynamic changes in histone modifications[1]. We show that SMCX is a novel histone demethylase with a substrate preference for H3K4me2 and H3K4me3 *in vitro* and *in vivo*. One of the DNA-binding proteins through which SMCX exerts its repressive effects is REST. The combined presence of REST, G9a, HDACs and SMCX in a repressive complex is consistent with previous reports indicating that REST mediates repression in part through G9a[9], and explains the simultaneous enrichment of H3K9me2 and depletion of H3K4me3 observed at REST target genes. Although mutations in REST have not yet been linked to neurological disorders, mutations or dysregulated expression of the SMCX/REST-regulated genes *SCG10*, *BDNF* and *SCN2A* are implicated in mental retardation and epilepsy[20,28,29]. Our identification of SMCX as a transcriptional regulator of several REST target genes is not only important for understanding SMCX-associated mental retardation, but will also provide greater insights into the molecular mechanisms underlying diverse cognitive disorders.

Note added in proof: While this manuscript was being reviewed, several groups reported that members of the JARID1 family possess H3K4 demethylase activity, confirming our finding that SMCX/JARID1C is an H3K4 tri- and di-demethylase[30–37].

METHODS SUMMARY

Procedures for SMCX complex purification, glycerol gradient sedimentation, and reciprocal immunoprecipitation have been described in detail previously[6,38].

Histone demethylase assay. Bulk histones, nucleosomes or histone peptides were incubated with GST–SMCX or SMCX complex in histone demethylase buffer (50 mm Tris, pH 8.0, 50 mM NaCl, 2 mM ascorbic acid, 1 mM α-ketoglutarate, 50 μM $Fe(NH)_4(SO_4)_2$, 0.1 Units formaldehyde dehydrogenase (FDH) and 1 mM NAD^+ at 32 °C for 2–8 h. In a typical reaction, 6 μg of calf thymus histones or nucleosomes, or 0.25–1 μg of histone peptides are incubated in a total reaction volume of 100 μl. The reaction mixture was analysed by SDS–polyacrylamide gel electrophoresis (PAGE) and western blotting using methyl-specific antibodies or by MALDI-TOF mass spectrometry.

MALDI-TOF mass spectrometry. The procedure for analysing peptide methylation has already been described in detail elsewhere[2]. The samples were analysed by a MALDI-TOF MS (PerSeptive Biosystems Voyager-DE STR Biospectrometry Workstation) at the Dana Farber Cancer Institute Molecular Biology Core Facilities.

RT–PCR superarray. The human neuroscience Ion Channels and Transporters RT^2 Profiler PCR Array (APHS-036A, SuperArray Bio) was used to obtain the expression profile of 84 genes encoding neuronal ion channels and transporters in stable HeLa cell lines expressing SMCX shRNA or control shRNA. Complementary DNA synthesis and real-time PCR was carried out in accordance with the manufacturer's instructions. Data shown are the average of duplicate PCR experiments.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 December 2006; accepted 10 April 2007.
Published online 29 April 2007.

1. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293,** 1074–1080 (2001).
2. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119,** 941–953 (2004).
3. Jensen, L. R. *et al.* Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am. J. Hum. Genet.* **76,** 227–236 (2005).
4. Tzschach, A. *et al.* Novel JARID1C/SMCX mutations in patients with X-linked mental retardation. *Hum. Mutat.* **27,** 389–394 (2006).
5. Chong, J. A. *et al.* REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* **80,** 949–957 (1995).
6. Ogawa, H., Ishiguro, K., Gaubatz, S., Livingston, D. M. & Nakatani, Y. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science* **296,** 1132–1136 (2002).
7. Kurokawa, R. *et al.* Polarity-specific activities of retinoic acid receptors determined by a co-repressor. *Nature* **377,** 451–454 (1995).
8. Ballas, N. *et al.* Regulation of neuronal traits by a novel transcriptional complex. *Neuron* **31,** 353–365 (2001).
9. Roopra, A., Qazi, R., Schoenike, B., Daley, T. J. & Morrison, J. F. Localized domains of G9a-mediated histone methylation are required for silencing of neuronal genes. *Mol. Cell* **14,** 727–738 (2004).
10. Wang, H. *et al.* Role of histone H2A ubiquitination in *Polycomb* silencing. *Nature* **431,** 873–878 (2004).
11. Tsukada, Y. *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439,** 811–816 (2006).
12. Whetstine, J. R. *et al.* Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125,** 467–481 (2006).
13. Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nature Rev. Genet.* **7,** 715–727 (2006).
14. Clissold, P. M. & Ponting, C. P. JmjC: cupin metalloenzyme-like domains in jumonji, hairless and phospholipase A2beta. *Trends Biochem. Sci.* **26,** 7–9 (2001).
15. Shi, Y. J. *et al.* Regulation of LSD1 histone demethylase activity by its associated factors. *Mol. Cell* **19,** 857–864 (2005).
16. Mori, N., Schoenherr, C., Vandenbergh, D. J. & Anderson, D. J. A common silencer element in the SCG10 and type II Na^+ channel genes binds a factor present in nonneuronal cells but not in neuronal cells. *Neuron* **9,** 45–54 (1992).
17. Mortazavi, A., Thompson, E. C., Garcia, S. T., Myers, R. M. & Wold, B. Comparative genomics modeling of the NRSF/REST repressor network: from single conserved sites to genome-wide repertoire. *Genome Res.* **16,** 1208–1221 (2006).
18. Vigo, E. *et al.* CDC25A phosphatase is a target of E2F and is required for efficient E2F-induced S phase. *Mol. Cell. Biol.* **19,** 6379–6395 (1999).
19. Kearney, J. A. *et al.* A gain-of-function mutation in the sodium channel gene Scn2a results in seizures and behavioral abnormalities. *Neuroscience* **102,** 307–317 (2001).
20. Lossin, C., Wang, D. W., Rhodes, T. H., Vanoye, C. G. & George, A. L. Jr. Molecular basis of an inherited epilepsy. *Neuron* **34,** 877–884 (2002).
21. Lubec, G. & Sohn, S. Y. RNA microarray analysis of channels and transporters in normal and fetal Down syndrome (trisomy 21) brain. *J. Neural Transm. Suppl.* **67,** 215–224 (2003).
22. Schousboe, A., Larsson, O. M., Sarup, A. & White, H. S. Role of the betaine/GABA transporter (BGT-1/GAT2) for the control of epilepsy. *Eur. J. Pharmacol.* **500,** 281–287 (2004).
23. Abdolmaleky, H. M., Thiagalingam, S. & Wilcox, M. Genetics and epigenetics in major psychiatric disorders: dilemmas, achievements, applications, and future scope. *Am. J. Pharmacogenomics* **5,** 149–160 (2005).
24. Lohoff, F. W. *et al.* Variations in the vesicular monoamine transporter 1 gene (VMAT1/SLC18A1) are associated with bipolar i disorder. *Neuropsychopharmacology* **31,** 2739–2747 (2006).
25. Splawski, I. *et al.* CACNA1H mutations in autism spectrum disorders. *J. Biol. Chem.* **281,** 22085–22091 (2006).
26. Johnson, R. *et al.* Identification of the REST regulon reveals extensive transposable element-mediated binding site duplication. *Nucleic Acids Res.* **34,** 3862–3877 (2006).
27. Kuwahara, K., Takano, M. & Nakao, K. Pathophysiological significance of T-type Ca^{2+} channels: transcriptional regulation of T-type Ca^{2+} channel–regulation of

CACNA1H by neuron-restrictive silencer factor. *J. Pharmacol. Sci.* **99,** 211–213 (2005).
28. Nelson, K. B. *et al.* Neuropeptides and neurotrophins in neonatal blood of children with autism or mental retardation. *Ann. Neurol.* **49,** 597–606 (2001).
29. Bahn, S. *et al.* Neuronal target genes of the neuron-restrictive silencer factor in neurospheres derived from fetuses with Down's syndrome: a gene expression study. *Lancet* **359,** 310–315 (2002).
30. Christensen, J. *et al.* RBP2 belongs to a family of demethylases, specific for tri-and dimethylated lysine 4 on histone 3. *Cell* **128,** 1063–1076 (2007).
31. Iwase, S. *et al.* The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. *Cell* **128,** 1077–1088 (2007).
32. Klose, R. J. *et al.* The retinoblastoma binding protein RBP2 is an H3K4 demethylase. *Cell* **128,** 889–900 (2007).
33. Lee, M. G., Norman, J., Shilatifard, A. & Shiekhattar, R. Physical and functional association of a trimethyl H3K4 demethylase and Ring6a/MBLR, a polycomb-like protein. *Cell* **128,** 877–887 (2007).
34. Lee, N. *et al.* The trithorax-group protein Lid is a histone H3 trimethyl-Lys4 demethylase. *Nature Struct. Mol. Biol.* (2007).
35. Secombe, J., Li, L., Carlos, L. & Eisenman, R. N. The Trithorax group protein Lid is a trimethyl histone H3K4 demethylase required for dMyc-induced cell growth. *Genes Dev.* **21,** 537–551 (2007).
36. Seward, D. J. *et al.* Demethylation of trimethylated histone H3 Lys4 *in vivo* by JARID1 JmjC proteins. *Nat. Struct. Mol. Biol.* **14,** 240–242 (2007).
37. Yamane, K. *et al.* PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. *Mol. Cell* **25,** 801–812 (2007).
38. Shi, Y. *et al.* Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* **422,** 735–738 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Shi and F. Lan for the gift of the His-JMJD2C protein. We also thank G. Mandel, T. Westbrook and P. Mulligan for sharing the anti-REST antibody and Y. Nakatani for the anti-RING2 antibody. This work was supported by NIH grants to Y.S. and by NIH and Harvard Stem Cell Institute (HSCI) grants to A.R. Y.S. is a PEW scholar. M.T. is supported by the HSCI and is a pre-doctoral fellow of the Ryan Foundation.

Author Contributions M.T. and P.M. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.S. (yujiang_shi@hms.harvard.edu).

METHODS

Recombinant protein expression and purification. GST–pBlueBac vector was obtained by inserting a GST tag into *Nhe*I and *Xho*1 sites of pBlueBac vector (Invitrogen). Wild-type SMCX was inserted into *Xba*I and *Hind*III sites of GST-pBlueBac. All point mutations were generated by site directed mutagenesis using the QuikChange Mutagenesis Kit (Stratagene). The sequences of all PCR-amplified clones were confirmed by conventional DNA sequencing. Baculovirus was generated using standard protocols (Invitrogen). Sf9 cells were infected with baculovirus for 2–3 days. The resulting cell pellet was kept on ice for 30 min in BC500-10 (20 mM Tris, pH 7.9, 500 mM NaCl, 1.5 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 10% glycerol, 0.1% Triton-X-100) with protease inhibitors, and then centrifuged at 10,000*g* for 20 min. The supernatant was adjusted to 300 mM NaCl by the addition of an equal volume of BC100-10 (20 mM Tris, pH 7.9, 100 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT, 10% glycerol, 0.1% Triton-X-100) and then incubated with GST affinity beads for 2 h at 4 °C. After extensive washing with BC100-10, the protein was eluted with 10 mM reduced glutathione in 20 mM Tris, pH 7.9, 50 mM NaCl. The homogeneity of the eluted protein was determined using SDS–PAGE followed by Coomassie blue staining.

Mass spectrometry. Flag-HA double-purified material was separated by 4–20% gradient SDS–PAGE and stained with colloidal Coomassie. Protein bands were excised, digested with tryspin and analysed by MS/MS mass spectrometry for peptide fragmentation at the Harvard Medical School Taplin Biological Mass Spectrometry Facility.

Immunofluorescence microscopy. U2OS cells were cultured for 24–48 h post-transfection, fixed with 3% paraformaldehyde (PFA), permeabilized in 0.1% Triton-X-100/phosphate buffered sulphate (PBS) and then blocked using 1% BSA/PBS. Coverslips were then incubated with primary antibodies diluted in 0.1% BSA/PBS for 1 h. After washing, coverslips were incubated in secondary antibody diluted in 0.1% BSA/PBS for 1 h. Coverslips were washed, stained with DAPI, mounted and analysed by fluorescence microscopy. Primary antibodies were used at the final dilutions: 1:800 anti-monomethyl-H3K4 (Upstate 07-436), 1:4,000 anti-dimethyl-H3K4 (Upstate 07-030), 1:1,000 anti-trimethyl-H3K4 (Upstate 07-473), 1:1,000 anti-HA (Sigma). Secondary antibodies used include goat anti-rabbit and goat-anti-mouse (1:1,000; Molecular Probes Alex 594 and Alexafluor 488).

Generation of stable SMCX knockdown cell lines. The pBABE-SMCX shRNA construct was generated against 5′-gcagcagaacatttggaagaa-3′, and the control shRNA construct was against 5′-gactcatcactcgataactga-3′. HeLa cells transduced with concentrated virus were selected with puromycin 24 h after transduction and were maintained in puromycin-containing media.

RNA interference and RT–PCR. HeLa cells were transfected with 100 pmol of siRNA duplexes using lipofectamine 2000 (Invitrogen). The duplexes against SMCX were synthesized by Invitrogen's Stealth-3 (5′-caagagacugcagccugauccg-gaa-3′). The efficiency of RNAi knockdown was determined by RT–PCR. Total RNA was extracted 42 h after transfection using Trizol (Life Science). RT–PCR was conducted using SuperScriptIII RT kit (Invitrogen) according to the manufacturer's instructions. The amplification efficiency of each set of primers was obtained by standard curve analysis. PCR products were separated by 2% agarose gel electrophoresis and visualized by ethidium bromide staining. Primer sequences are available upon request.

ChIP and sequential ChIP. Cells were cross-linked with 1% formaldehyde for 10 min at room temperature. Cross-linking was stopped by the addition of glycine to 0.125 M. Cells were washed three times in ice-cold 1 × PBS and resuspended in lysis buffer (0.2% SDS, 10 mM EDTA, 50 mM Tris, pH 8.0, and protease inhibitors) and was sonicated to obtain DNA fragments with an average size of 500 base pairs. Sonicated chromatin was mixed with one volume of TE buffer (10 mM Tris, pH 8.0, 1 mM EDTA) and two volumes of 2 × ChIP binding buffer (50 mM Tris, pH 8.0, 300 mM NaCl, 2 mM EDTA and 1% Triton X-100). For anti-Flag ChIP, chromatin from 3×10^7 cells was incubated overnight with 80 µl anti-Flag M2 affinity gel (Sigma, A2220) preincubated with 50 µg sheared salmon sperm DNA and 100 µg of bovine serum albumin (BSA). As a control, 0.2 mg ml^{-1} flag peptide (Sigma F3290) was added to block the binding of the anti-Flag antibody. Beads were washed three times with low salt wash buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, and 1% Triton X-100), once with high-salt wash buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 2 mM EDTA, 0.1% SDS, and 1% Triton X-100) and once with TE. Beads were eluted twice for 30 min at room temperature using 200 µl 0.2 mg ml^{-1} Flag peptide in 1 × ChIP binding buffer. Flag eluate was used to extract DNA for quantitative PCR or used as the input for a tandem immunoprecipitation.

For ChIP using anti-REST antibody (Upstate 07-579), anti-H3K4me3 antibody (Abcam 8580-100), anti-H3K4me2 antibody (Upstate 07-030), anti-H3K9me2 (Upstate 07-441) and anti-H3-acetyl antibody (Upstate 06-599), chromatin from 0.4–1 × 10^7 cells was precleared with a mixture of protein A/G conjugated agarose (Upstate 16-257 and 16-201C) preincubated with salmon sperm DNA and BSA and incubated overnight with 1 µg antibody at 4 °C. Sequential ChIP was performed by incubating anti-Flag immunoprecipitated material with 1 µg of the antibody of interest or with control IgG antibody overnight at 4 °C. Immunoprecipitated complexes were recovered by 60 µl of protein A-protein G conjugated agarose, washed as described, eluted with 1% SDS and 0.1 M $NaHCO_3$ and de-crosslinked at 65 °C for 5 h. DNA was purified using Qiagen MinElute PCR Purification kit (28006). The pulldown efficiency of each gene target was determined using quantitative real-time PCR (Bio-Rad IQTM SYBR Green Supermix 170-8882).

The amplification efficiency of each set of primers was obtained by standard curve analysis. ChIP primer sequences are as follows: SCG10-RE-F (5′-ccagtag-catcctatcagt-3′), SCG10-RE-R (5′-cctcaggaacaaacacag-3′), SCG10-2-F (5′-ttag-gactcaatcgtgagg-3′), SCG10-2R (5′-ctccacccatctttcaaac-3′), BDNF-RE-F (5′-agagccaacggatttgtc-3′), BDN-FRE-R (5′-cttgccaagagtctattcc-3′), CDC25A-F (5′-gacctccttcaagtacaaag-3′), CDC25A-R (5′-tggacacagatgactgatg-3), POLII-EX-F (5′-gctcctatgtcaattaccg-3′), POLII-EX-R (5′-gttaccgtttcctcaaagg-3′), SCN2A-2F (5′-cgtgtttcaaggctacagca-3′), SCN2A-2R (5′-ctctagcctcccaaccttcc-3′), SCN2A-RE-F (5′-tcaattattctggctgatctgg-3′), SCN2A-RE-R (5′-ggaaattaaa-gatcgatttggc-3′), SYN1-RE-F, (5′-tgggttttaggaccaggatg-3′), SYN1-RE-R, (5′-ggtgctgaagctggcagt-3′).

Vol 447|31 May 2007|doi:10.1038/nature05843

LETTERS

8-oxo-guanine bypass by human DNA polymerases in the presence of auxiliary proteins

Giovanni Maga[1,4], Giuseppe Villani[2], Emmanuele Crespan[1], Ursula Wimmer[4], Elena Ferrari[4], Barbara Bertocci[3] & Ulrich Hübscher[4]

Specialized DNA polymerases (DNA pols) are required for lesion bypass in human cells[1]. Auxiliary factors have an important, but so far poorly understood, role. Here we analyse the effects of human proliferating cell nuclear antigen (PCNA) and replication protein A (RP-A) on six different human DNA pols—belonging to the B, Y and X classes—during *in vitro* bypass of different lesions. The mutagenic lesion 8-oxo-guanine (8-oxo-G) has high miscoding potential[2–4]. A major and specific effect was found for 8-oxo-G bypass with DNA pols λ and η. PCNA and RP-A allowed correct incorporation of dCTP opposite a 8-oxo-G template 1,200-fold more efficiently than the incorrect dATP by DNA pol λ, and 68-fold by DNA pol η, respectively. Experiments with DNA-pol-λ-null cell extracts suggested an important role for DNA pol λ. On the other hand, DNA pol ι, together with DNA pols α, δ and β, showed a much lower correct bypass efficiency. Our findings show the existence of an accurate mechanism to reduce the deleterious consequences of oxidative damage and, in addition, point to an important role for PCNA and RP-A in determining a functional hierarchy among different DNA pols in lesion bypass.

Our experiments show that DNA pol λ inserted either dATP or dCTP *in vitro* opposite an 8-oxo-G lesion on a template DNA oligonucleotide, and, unlike pol μ (ref. 5), did so without −1 frameshifts (Supplementary Fig. 2a), and showing a 12.5-fold higher incorporation efficiency (k_{cat}/K_m) for faithful dCTP versus error-prone dATP incorporation (Supplementary Table 1). DNA pol λ physically and functionally interacts with the proteins PCNA and RP-A (refs 6–8).

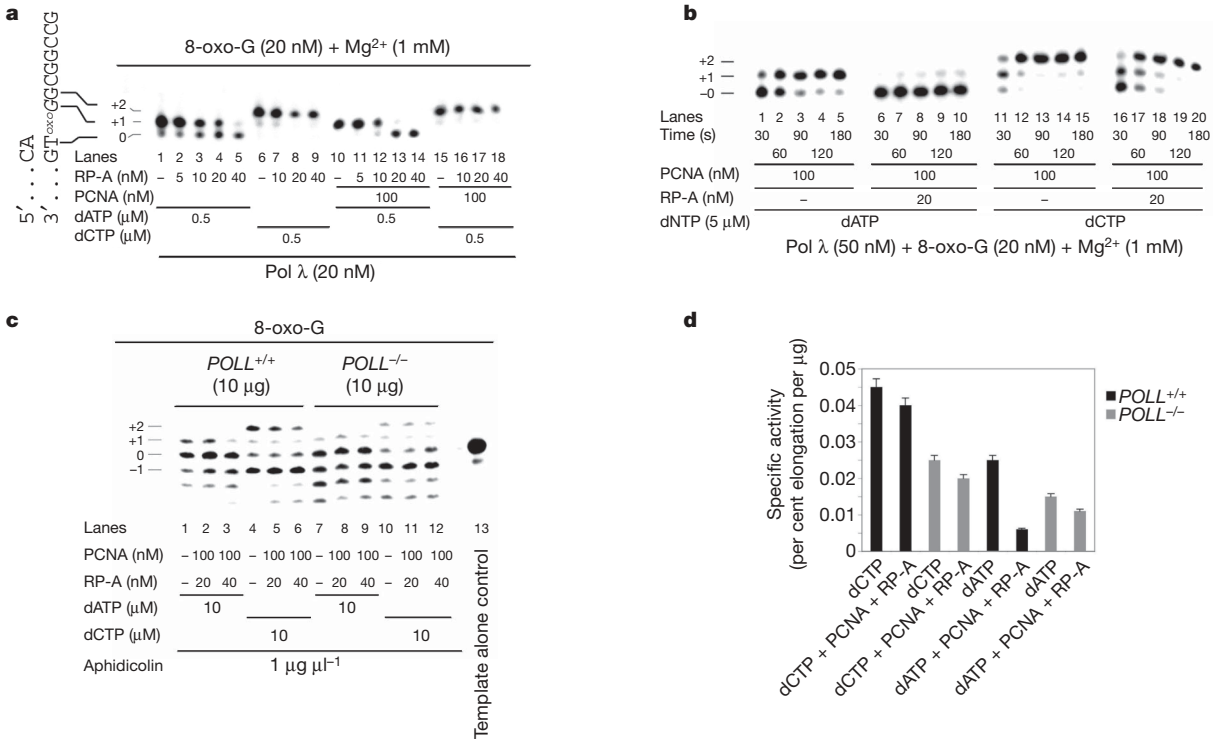


Figure 1 | PCNA and RP-A influence bypass of 8-oxo-G by human DNA pol λ. **a**, dATP (lanes 1–5 and 10–14) or dCTP (lanes 6–9 and 15–18) incorporation opposite 8-oxo-G with RP-A or PCNA. **b**, Time-dependent dATP or dCTP incorporation opposite 8-oxo-G with or without PCNA and RP-A. **c**, 8-oxo-G bypass by $POLL^{+/+}$ or $POLL^{-/-}$ MEF extracts with PCNA, RP-A and dATP (lanes 1–3 and 7–9) or dCTP (lanes 4–6 and 10–12). **d**, Relative dATP or dCTP incorporation opposite 8-oxo-G with or without PCNA and RP-A by $POLL^{+/+}$ (black) or $POLL^{-/-}$ (grey) MEF extracts. Values are means of three independent replicates ± s.d.

[1]Institute of Molecular Genetics IGM-CNR, via Abbiategrasso 207, I-27100 Pavia, Italy. [2]Institute de Pharmacologie et de Biologie Structurale, IPBS-CNRS, Centre National de la Recherche Scientifique, Route de Narbonne 205, 31077 Toulouse, France. [3]Institute National Français de Recherche Médicale, Université Paris Descartes, Faculté de Médecine René Descartes, Site Necker-Enfants Malades, 156 rue de Vaugirard 75730 Paris, Cedex 15, France. [4]Institute for Veterinary Biochemistry and Molecular Biology, University of Zürich-Irchel, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.

PCNA increased the incorporation rates (k_{obs}) of dCTP opposite a normal G and 8-oxo-G, but not of dATP incorporation opposite the lesion (Supplementary Fig. 2b–d). dATP (Fig. 1a, lanes 1–5) but not dCTP (lanes 6–9) incorporation opposite 8-oxo-G was inhibited by RP-A alone and, even more efficiently, by RP-A and PCNA together (Fig. 1a, lanes 13 and 17). In the presence of PCNA, addition of RP-A reduced dATP incorporation opposite 8-oxo-G to background levels (Fig. 1b compare lanes 1–5 with lanes 6–10), whereas dCTP incorporation was affected less than twofold (Fig. 1b lanes 11–15 and 16–20). In the presence of PCNA and RP-A, DNA pol λ catalysed only 1 error-prone dATP incorporation every 1.2×10^3 faithful dCTP incorporation events opposite 8-oxo-G (Table 1 and Supplementary Fig. 2e), which is sufficient to efficiently cope with the $\sim 10^3$ lesions of 8-oxo-G estimated to be present in normal cells[2]. From the difference in the k_{cat}/K_m values for dATP versus dCTP incorporation in the presence of PCNA (Supplementary Table 1), a relative free energy change ($\Delta\Delta G$) value of $+1.9$ kcal mol^{-1} was estimated to be required for generation of an A:8-oxo-G mismatch by DNA pol λ. Such a decreased thermodynamic stability may facilitate displacement of DNA pol λ by RP-A when it is incorporating dATP opposite 8-oxo-G. Identical results were obtained with two unrelated sequence contexts (Supplementary Fig. 2f, g). BSA or *Escherichia coli* single strand DNA binding protein could not substitute for RP-A (data not shown and refs 6, 8).

Next, extracts of mouse embryonic fibroblasts (MEFs) with or without DNA pol λ (*POLL*$^{+/+}$ and *POLL*$^{-/-}$, respectively) were used for *in vitro* 8-oxo-G bypass reactions. Aphidicolin was added to reduce background error-prone synthesis by DNA pols α, δ, ε and ζ. With the *POLL*$^{+/+}$ extract, the ratio of dCTP versus dATP incorporation increased from 1.8 in the absence of PCNA and RP-A to 6.6 in their presence, contrary to the *POLL*$^{-/-}$ extract (1.5 and 1.8 ratio values, respectively, Fig. 1c, d). The endogenous concentrations of PCNA and RP-A were too low to contribute to the observed effects (1 μg of extract corresponded to a 1 nM final concentration of PCNA in the assay, Supplementary Fig. 3a). Addition of recombinant DNA pol λ to the *POLL*$^{-/-}$ extract restored the stimulation of the dCTP versus dATP incorporation by PCNA and RP-A (Supplementary Fig. 3b), suggesting a specific role of DNA pol λ in faithful 8-oxo-G bypass together with PCNA and RP-A.

Next we tested other DNA pols in the same experimental system. Human DNA pol η incorporated dCTP opposite an 8-oxo-G with the same efficiency (k_{cat}/K_m) as opposite a normal G (Supplementary Table 1)[9], showing a 2.5-fold preference for dCTP versus dATP incorporation opposite an 8-oxo-G (Table 1). RP-A and PCNA, selectively suppressed dATP versus dCTP incorporation opposite an 8-oxo-G (Fig. 2 a–c), raising the preference for faithful versus error-prone 8-oxo-G bypass by DNA pol η to 68-fold (Table 1). This makes DNA pol η the second most accurate enzyme after DNA pol λ at 8-oxo-G.

DNA pol ι incorporates dGTP with similar efficiency to dCTP opposite an 8-oxo-G (Supplementary Table 1)[10]. PCNA stimulated dCTP incorporation opposite the lesion (Fig. 2d, compare lanes 2 and 3 with lanes 12 and 13), but no further elongation was observed[10], whereas RP-A completely inhibited 8-oxo-G bypass (lanes 5–8, 15–18). PCNA restored the ability of DNA pol ι to bypass the 8-oxo-G lesion in the presence of RP-A, resulting in a 5-fold preference for dCTP (lanes 9, 10) versus dGTP (lanes 19, 20) incorporation (Table 1). Thus, PCNA is important for 8-oxo-G lesion bypass by DNA pol ι in the presence of RP-A. DNA pol α does not interact with PCNA, but

Table 1 | The preference for faithful dCTP incorporation opposite 8-oxo-G by human DNA polymerase λ in the presence of PCNA and RP-A can allow accurate bypass of the 'physiological' levels of this oxidative lesion*

DNA polymerase (DNA pol) and proteins	Family	Function	dCTP versus dATP† or dGTP‡
DNA pol α	B	Replication	1:10
DNA pol α, RP-A		Replication	1:2
DNA pol δ	B	Replication	1:2
DNA pol δ, PCNA		Replication	3:1
DNA pol δ, PCNA, RP-A		Replication	3:1
DNA pol β	X	Repair	2:1
DNA pol β, PCNA		Repair	2:1
DNA pol β, RP-A		Repair	2:1
DNA pol β, PCNA, RP-A		Repair	3:1
DNA pol η	Y	Translesion	2.5:1
DNA pol η, PCNA		Translesion	20:1
DNA pol η, PCNA, RP-A		Translesion	68:1
DNA pol ι	Y	Translesion	2:1‡
DNA pol ι, PCNA, RP-A		Translesion	5:1‡
DNA pol λ	X	Repair	12:1
DNA pol λ, PCNA		Repair	24:1
DNA pol λ, PCNA, RP-A		Repair	1200:1

* The steady state level of 8-oxo-G as measured by high-performance liquid chromatography or enzymatic conversion of strand breaks is estimated to be ~1,000–10,000 per cell (see ref. 2 for details and citations).
† Derived from the selectivity (1/f) values determined in this study. See Supplementary Table 1 and Methods for details.
‡ Bias dCTP versus dGTP incorporation.

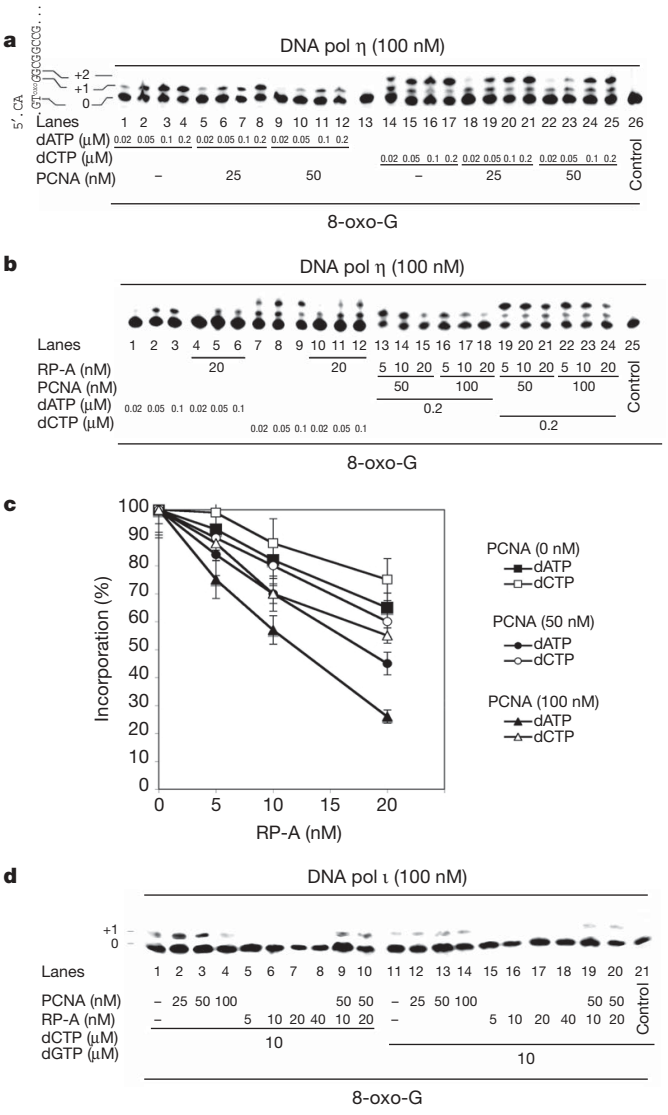


Figure 2 | 8-oxo-G bypass by human DNA pols η and ι with RP-A and PCNA. **a**, dATP or dCTP incorporation by DNA pol η with PCNA. **b**, dATP (lanes 1–6 and 13–18) or dCTP (lanes 7–12 and 19–24) incorporation by DNA pol η with RP-A (lanes 4–6 and 10–12) or RP-A and PCNA (lanes 13–24). **c**, dATP (black symbols) or dCTP (open symbols) incorporation by human DNA pol η with RP-A, or RP-A and PCNA. Error bars, ±s.d. of three independent replicates. **d**, dCTP (lanes 1–10) or dGTP (lanes 11–21) incorporation by DNA pol ι with PCNA (lanes 2–4 and 12–14), RP-A (lanes 5–8 and 15–18) or both (lanes 9, 10, 19 and 20).

its fidelity is increased by RP-A[11]. As shown in Supplementary Fig. 4a, b, RP-A selectively inhibited dATP more than dCTP incorporation by DNA pol α opposite an 8-oxo-G (Supplementary Table 1). However, lesion bypass was still highly error-prone even in the presence of RP-A (Table 1). Finally, PCNA and RP-A under similar conditions did not affect the bypass of 8-oxo-G by DNA pols δ and β (Table 1 and Supplementary Fig. 4c, d).

DNA pol λ bypasses an abasic site exclusively by template-slippage, generating −1 frameshifts, whereas pol β can either skip the lesion, or incorporate dATP (ref. 12). PCNA and RP-A have been shown to promote bypass of an abasic site by Y-family DNA pols[13,14], and PCNA alone stimulated abasic site bypass by DNA pol λ (ref. 7; Supplementary Fig. 5a). In the presence of PCNA, DNA pol λ bypassed the abasic site by inserting dCTP opposite the G on the template position downstream of the lesion (Supplementary Fig. 5b, lanes 1–4) but could not incorporate dATP (Supplementary Fig. 5b, lanes 5–10). RP-A did not influence the abasic-site bypass by DNA pol λ, either in the absence or in the presence of PCNA (Supplementary Fig. 5b, lanes 12–20). Also, PCNA and RP-A did not show any effect on abasic-site bypass by DNA pol β (Supplementary Fig. 5c).

The GG *cis*-platinum adduct has been shown to block synthesis by DNA pols α, δ and λ, even in the presence of PCNA[7,15], whereas DNA pols β and η can bypass this lesion[15,16]. PCNA and RP-A did not significantly influence the bypass of a single *cis*-platinum adduct by DNA pol β (Supplementary Fig. 5d) or DNA pol η (Supplementary Fig. 5e), suggesting that bypass of bulky adducts is primarily dependent on the intrinsic properties of the DNA pol involved.

In summary our results suggest a specific role of RP-A and PCNA in the 8-oxo-G lesion bypass by DNA pol λ and—to a lower extent—by DNA pol η, and indicate that the effects of these two auxiliary proteins on the bypass activity of different DNA pols can be modulated by the nature of the lesions.

METHODS SUMMARY

Nucleic acid substrates. All the oligonucleotide templates, either undamaged or containing the lesions, have been chemically synthesized and purified from polyacrylamide gels. Concentrations were determined spectrophotometrically. For sequences and details see Methods.

Proteins production and purification. All recombinant human proteins were expressed and purified according to published procedures (see Methods).

Enzymatic assays. *In vitro* assays were performed as described in Methods. Products were resolved on a 7 M urea/10% polyacrylamide gel and visualized by PhosphoImager (Typhoon Trio GE Healthcare).

Cells and extracts. Extracts were prepared from 10^7 immortalized *POLL*$^{+/+}$ and *POLL*$^{-/-}$ MEFs in the presence of 0.4 M NaCl. For full details see Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 November 2006; accepted 22 March 2007.
Published online 16 May 2007.

1. Hubscher, U., Maga, G. & Spadari, S. Eukaryotic DNA polymerases. *Annu. Rev. Biochem.* **71,** 133–163 (2002).
2. Collins, A. R. Oxidative DNA damage, antioxidants, and cancer. *Bioessays* **21,** 238–246 (1999).
3. Krahn, J. M., Beard, W. A., Miller, H., Grollman, A. P. & Wilson, S. H. Structure of DNA polymerase β with the mutagenic DNA lesion 8-oxodeoxyguanine reveals structural insights into its coding potential. *Structure* **11,** 121–127 (2003).
4. Shibutani, S., Takeshita, M. & Grollman, A. P. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature* **349,** 431–434 (1991).
5. Zhang, Y. *et al.* Lesion bypass activities of human DNA polymerase μ. *J. Biol. Chem.* **277,** 44582–44587 (2002).
6. Maga, G., Shevelev, I., Villani, G., Spadari, S. & Hubscher, U. Human replication protein A can suppress the intrinsic *in vitro* mutator phenotype of human DNA polymerase λ. *Nucleic Acids Res.* **34,** 1405–1415 (2006).
7. Maga, G. *et al.* Human DNA polymerase λ functionally and physically interacts with proliferating cell nuclear antigen in normal and translesion DNA synthesis. *J. Biol. Chem.* **277,** 48434–48440 (2002).
8. Maga, G. *et al.* DNA elongation by the human DNA polymerase λ polymerase and terminal transferase activities are differentially coordinated by proliferating cell nuclear antigen and replication protein A. *J. Biol. Chem.* **280,** 1971–1981 (2005).
9. Haracska, L., Yu, S. L., Johnson, R. E., Prakash, L. & Prakash, S. Efficient and accurate replication in the presence of 7,8-dihydro-8-oxoguanine by DNA polymerase η. *Nature Genet.* **25,** 458–461 (2000).
10. Vaisman, A. & Woodgate, R. Unique misinsertion specificity of pol ι may decrease the mutagenic potential of deaminated cytosines. *EMBO J.* **20,** 6520–6529 (2001).
11. Maga, G., Frouin, I., Spadari, S. & Hubscher, U. Replication protein A as a "fidelity clamp" for DNA polymerase α. *J. Biol. Chem.* **276,** 18235–18242 (2001).
12. Blanca, G. *et al.* Human DNA polymerases λ and β show different efficiencies of translesion DNA synthesis past abasic sites and alternative mechanisms for frameshift generation. *Biochemistry* **43,** 11605–11615 (2004).
13. Haracska, L., Kondratick, C. M., Unk, I., Prakash, S. & Prakash, L. Interaction with PCNA is essential for yeast DNA polymerase η function. *Mol. Cell* **8,** 407–415 (2001).
14. Haracska, L. *et al.* Targeting of human DNA polymerase ι to the replication machinery via interaction with PCNA. *Proc. Natl Acad. Sci. USA* **98,** 14256–14261 (2001).
15. Hoffmann, J. S. *et al.* DNA polymerase β bypasses *in vitro* a single d(GpG)-cisplatin adduct placed on codon 13 of the *HRAS* gene. *Proc. Natl Acad. Sci. USA* **92,** 5356–5360 (1995).
16. Bassett, E. *et al.* Frameshifts and deletions during *in vitro* translesion synthesis past Pt-DNA adducts by DNA polymerases β and η. *DNA Repair (Amst.)* **1,** 1003–1016 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Woodgate for his generous gift of recombinant human DNA pol η and ι. U.H. and U.W. are supported by the Swiss National Science Foundation, by the UBS "im Auftrag eines Kunden", and U.H. and E.F. by the University of Zürich, which gave a grant in aid to G.M. G.M. is supported partially by the CARIPLO Foundation Project "Oncogenetica e Proteomica della Replicazione". G.V. is supported by CNRS and ARC.

Author Contributions G.M. and U.H. had the original idea. G.M. supervised the overall experimental strategy and performed all the experiments with human DNA pol λ and cell extracts; E.C. performed all the experiments with human DNA polymerases α, δ, β, ι and η; U.W. and B.B. generated the DNA *POLL*$^{+/+}$ and *POLL*$^{-/-}$ MEFs and characterized their phenotype; G.V. provided the damaged templates; E.F. purified DNA pols α, δ, PCNA and RP-A; G.M., G.V. and U.H. designed and interpreted all the experiments and equally contributed to manuscript writing and figures preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.M. (maga@igm.cnr.it).

METHODS

Chemicals. [γ-^{32}P]ATP (3,000 Ci mmol^{-1}) was from GE Healthcare Biosciences; unlabelled dNTPs were from Roche Molecular Biochemicals. All other reagents were of analytical grade and were purchased from Merck or Fluka.

Nucleic acids substrates. The sequences are:

72-mer 8-oxo-G template, 3′ATGTTGGTTCTCGTATGCTGCCGGTCACG-GCTTAAGTGT**G**GCGGCCGCGGGTTGGAGGGCTTATAGATTATG;

73-mer abasic template, 3′ATAGGTGGTATGATGGGAGGTGATAGAGGT-GAGTTGAGTTGGGAAGTA**G**GAGTTATAAGGATGGGAGGGCTAG;

60-mer *cis*-platinum template, 3′ATGTTGGTTCTCGTATGCTGCCGGTCA-CGGCTTTTCTT**GG**TTCCTATCGGTGGTTAGTCG.

The 72-mer and 73-mer, either undamaged or containing the 8-oxo G or abasic lesions (8-oxo-G-CE Phosphoramidite, or Tetrahydrofuran from Glen Research), and the corresponding primer were chemically synthesized. The 60-mer template containing a single *cis*-platinum adduct was prepared as described[15]. All substrate oligonucleotides and the corresponding primers were purified on a 15% (w/v) polyacrylamide, 7 M urea, 30% formamide gel. After elution and ethanol precipitation, their concentrations were determined by spectrophotometry. The bold letters correspond to the positions of the 8-oxo-G, abasic or *cis*-platinum lesions, respectively. The underlined sequences correspond to the primer annealing sites. The primer was 5′-labelled with T4 polynucleotide kinase (New England Biolabs) in the presence of [γ-^{32}P]ATP. Each labelled primer was mixed with the complementary template oligonucleotide at 1:1 (M/M) ratio in the presence of 25 mM Tris-HCl (pH 8.0) and 50 mM KCl, heated at 80 °C for 3 min and then slowly cooled down at to 20 °C.

Proteins production and purification. Recombinant His-tagged human wild-type DNA pol λ, human PCNA and human RP-A were expressed and purified as described[6]. Human DNA pol α and human DNA pol δ were purified from HeLa cells according to the same procedure previously established for calf thymus[17]. Human DNA pol β was from TREVIGEN. After purification, the proteins were >90% homogenous, as judged by SDS–PAGE and Coomassie staining.

Cells and extracts.. Primary *POLL*$^{+/+}$ and *POLL*$^{-/-}$ MEFs were generated as previously described[18]. Immortalized MEFs were obtained after 15 passages with a 3-day transfer regime according to a standard 3T3 culture protocol[19]. The cells were grown at 37 °C in a 5% CO_2 incubator in DMEM containing GlutaMAX-I (Gibco BRL), 10% fetal calf serum (Brunschwig), and 100 U ml^{-1} penicillin–streptomycin (Gibco BRL). Extracts were prepared by adding 0.5 ml of lysis buffer (10 mM TrisHCl (pH 7.5) 1 mM EDTA (pH 8.0), 10% sucrose, 0.4 M NaCl, 1 mM DTT, 0.1 mM PMSF, protease inhibitors) to 10^7 *POLL*$^{+/+}$ or *POLL*$^{-/-}$ MEFs. Cells were incubated in lysis buffer for 30 min on ice, homogenized by Dounce and centrifuged for 20 min at 4 °C at 15,000 × *g*. The supernatant (cell extract) was aliquoted and kept at −80 °C until use. Protein concentrations were 7 mg ml^{-1} for *POLL*$^{+/+}$ and 8.5 mg ml^{-1} for *POLL*$^{-/-}$ MEFs.

Enzymatic assays. For denaturing gel analysis of DNA synthesis products, the reaction mixtures contained 50 mM Tris-HCl (pH 7.0), 0.25 mg ml^{-1} BSA, 1 mM DTT and 20 nM (0.2 pmol of 3′ OH ends) of the 5′ ^{32}P-labelled primer/template (unless otherwise stated). DNA pols α, δ, η and ι were assayed in the presence of 5 mM Mg^{2+}, whereas for DNA pols λ and β a final Mg^{2+} concentration of 1 mM was used. Concentrations of DNA pols, PCNA, RP-A, dNTPs were as indicated. Reactions were incubated for 5 min at 37 °C, unless otherwise stated, and then stopped by addition of standard denaturing gel loading buffer (95% formamide, 10 mM EDTA, xylene cyanol and bromophenol blue), heated at 95 °C for 3 min and loaded on a 7 M urea/10% polyacrylamide gel.

Steady state kinetic analysis. Reactions were performed as described above. Quantification was done by scanning densitometry. The initial velocities of the reaction were calculated from the values of integrated gel band intensities:

$$I^*_T/I_{T\text{-}1}$$

where T is the target site, the template position of interest; I^*_T is the sum of the integrated intensities at positions T, $T + 1$, ..., $T + n$.

All the intensity values were normalized to the total intensity of the corresponding lane to correct for differences in gel loading. The apparent K_m and k_{cat} values were calculated by plotting the initial velocities in dependence of the nucleotide concentrations [dNTP] and fitting the data according to the Michaelis–Menten equation:

$$k_{cat}[E]_0/(1+K_m/[dNTP])$$

where $[E]_0$, was the input enzyme concentration. Nucleotide concentrations used (unless otherwise stated in the Figs) were 0.002 μM, 0.005 μM, 0.02 μM, 0.05 μM, 0.2 μM, 0.5 μM, 2 μM, 5 μM and 10 μM.

Nucleotide incorporation efficiencies were defined as the k_{cat}/K_m ratio. Under single nucleotide incorporation conditions $k_{cat} = k_{pol}k_{off}/(k_{pol} + k_{off})$ and $K_m = K_s k_{off}/(k_{pol} + k_{off})$, where k_{pol} is the true polymerization rate, k_{off} is the dissociation rate of the enzyme–primer complex and K_s is the true Michaelis constant for nucleotide binding. Thus, k_{cat}/K_m values are equal to k_{pol}/K_s.

Time-dependent accumulation of products was fitted to the mixed exponential equation:

$$[P] = A(1 - e^{-k_{obs}t}) + k_{ss}t$$

where [P] are the products formed, A is the burst amplitude, t is time, k_{obs} is the apparent burst rate and k_{ss} is the steady-state rate, which corresponds to k_{off}.

Free energy change ($\Delta\Delta G$) values were estimated according to the relationship:

$$\Delta G = \mathrm{RT}\ln(k)$$

where R is the gas constant (1.9872 cal mol^{-1} K^{-1}), T is the absolute temperature (K), and k is the ratio between the kinetic constants considered.

17. Weiser, T. *et al.* Biochemical and functional comparison of DNA polymerases α, δ and ε from calf thymus. *J. Biol. Chem.* **266**, 10420–10428 (1991).
18. Bertocci, B., De Smet, A., Weill, J. C. & Reynaud, C. A. Nonoverlapping functions of DNA polymerases μ, λ, and terminal deoxynucleotidyltransferase during immunoglobulin V(D)J recombination *in vivo*. *Immunity* **25**, 31–41 (2006).
19. Todaro, G. J. & Green, H. Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. *J. Cell Biol.* **17**, 299–313 (1963).

naturejobs

THE CAREERS MAGAZINE FOR SCIENTISTS

Two seemingly opposite definitions of the word 'career' came into play earlier this month at a round-table discussion of professional opportunities in science, sponsored by *Naturejobs* at the University of Birmingham, UK. "A course or progress through life...an occupation or profession engaged in as a life-work" are definitions in the *Oxford English Dictionary*, and the ones that probably first come to mind. They indicate a certain amount of planning and imply some sort of linear progression.

But Alison Cooper, a lecturer in pharmacology at the University of Birmingham, cited another definition that applied to all seven of us panellists: "Move or swerve about wildly."

The job in which one ends up is very rarely completely due to chance. But, like many discoveries in science, a certain amount of serendipity — and conditions in the real world — had come into play in our cases. Cooper originally intended to focus on *in vivo* pharmacology, but at the time she finished her PhD, there was little demand for those particular skills in industry; that, of course, has changed completely over the past decade. So she became an instructor.

Andy Garner, a group leader in drug discovery at AstraZeneca, applied for a job at the company and attended the interview just for practice — and landed a position.

And I intended to be a medical writer for newspapers, but quite literally by accident (see *Nature* **430,** 705; 2004) found myself teaching and carrying out research at the University of Wisconsin, Madison, before spending the past ten years writing for a scientific audience.

Despite this erratic 'careering', none of the panelists' current destinations was a complete accident. All had identified transferable skills, then broadened them, weathered storms of rejection, tapped into personal and professional networks, and, perhaps most important, sought the best person, place and project to fit their skills and interests. In doing so, we all met the first definition of 'career' as well.

Paul Smaglik, *Naturejobs* editor

CONTACTS
Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
***Nature*events:**
Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:
Muriel Lestringuez (4994)
Southwest UK/RoW:
Nils Moeller (4953)
Scandinavia/Spain/Portugal/Italy:
Evelina Rubio-Hakansson (4973)
Northeast UK/Ireland:
Matthew Ward (+44 (0) 20 7014 4059)
North Germany/The Netherlands:
Reya Silao (4970)
South Germany/Austria:
Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:
Stephen Russell
To send materials use London
address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:**
Tom Hancock

***Naturejobs* online production:**
Jasmine Myer
US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku,
Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

NATURE|Vol 447|31 May 2007

MOVERS

Carla Shatz, director, BioX, Stanford University, Stanford, California

2000–07: Professor and chair, Department of Neurobiology, Harvard Medical School, Boston, Massachusetts
1992–2000: Professor of neurobiology, University of California, Berkeley
1978–91: From assistant to full professor, Stanford University, Stanford, California

Carla Shatz has blazed a trail of firsts for women in neuroscience. Along the way, she has mentored a number of female scientists. She thinks that most girls are interested in science, but many simply lose interest under pressure from society.

After a chemistry degree from the Radcliffe Institute for Advanced Study at Harvard, she received a Marshall Fellowship to University College London and spent the next two years learning about physiology and biological systems.

Shatz returned to the United States to receive the first doctorate in neurobiology from Harvard — in part because she thought the move would offer broader opportunities. Being mentored by David Hubel and Torsten Wiesel, who were conducting Nobel-prizewinning science detailing how the brain is wired, made it an even smarter move.

"I got interested in this incredibly beautiful computational machine, and I wanted to see how the heck it was wired up during development," she says.

For her first job, Shatz's strategy was to go to the place that really wanted her. When Stanford University came calling, Shatz accepted. There, she became the first female basic scientist to be granted tenure at the medical school.

After 13 years at Stanford, she spent the next eight at the University of California, Berkeley, where she tried to establish an interdisciplinary neuroscience institute. Frustrated by a lack of administrative support, Shatz accepted an offer to be the first woman to chair Harvard's neurobiology department. "I couldn't turn it down because I felt I was on a mission to represent women at the highest levels of the university," she says. But she says the competition for space and resources were difficult to handle.

Having always hoped to return to Stanford, Shatz will soon be leading its BioX programme, which merges science, medicine and technology. Stanford neurobiologist William Newsome describes Shatz as an ideal leader for BioX because she has used a host of approaches — from anatomy to molecular biology to physiology — to show that brain development is not solely under genetic control. Shatz says that she sees her biggest challenge as integrating BioX with other departments.

Former colleagues eagerly await her return. "BioX already has fruit, but Carla will grow that tree and expand its influence," says Susan McConnell, a former graduate student of Shatz, adding that one of her many exceptional talents is facilitating long-lasting communities. ■

Virginia Gewin

MENTORS & PROTÉGÉS

Award-winning commitment

Walter Dodds, a biology professor at Kansas State University in Manhattan, likes to do research and to analyse data, but he likes encouraging his students even more. This year, the university's biology graduate student association presented Dodds with the Outstanding Graduate Faculty award. Among the criteria for this honour were teaching skills, mentorship, the cultivation of learning opportunities, participation in graduate student committees and publication and grants record.

Dodds teaches limnology and aquatic and microbial ecology, sciences that arouse his enthusiasm in the classroom and on field trips to nearby streams in the Konza Prairie. His contagious zeal makes learning come naturally.

Dodds says his goals are to "nurture a love for scientific research", and "make students successful in their career paths". Students find this fosters a passionate and professional attitude towards their work. As well as serving on almost every committee of students whose research relates to aquatic ecology, Dodds recently worked with other faculty members to create a course on professional skills in biology — a core requirement for incoming graduate students.

"His door is always open for us," says one of the many students who wrote in to support Dodds's nomination, adding that he "keeps in touch with his past students, meeting up for dinner at conferences".

Other supporters note that Dodds is always eager to work on field trips and lab demonstrations — even in cleaning up — and that he strives to end all his graduate courses with a publication that lists all students as authors. Quick to stem the tide of self-doubt that engulfs all graduate students at some point, he is understanding and flexible towards their family lives. During hours off, he invites students to the pub, or to enjoy his harmonica-playing in the locally renowned Red State Blues Band.

Dodds's professional accolades are impressive. His research has so far secured several million dollars in extramural support for the university, and his long list of publications documents a robust career of quality science. It was not his career, however, that prompted so many letters of support to the award committee. It was his intangible knack of mentorship, the ability to convey to students the idea that, in fact, they are his finest achievement. ■

Ron VanNimwegen is a PhD candidate in biology at Kansas State University.

POSTDOC JOURNAL

Loose ends

My PhD supervisor once told me that your thesis is never really completed. Given that his PhD, like mine, involved palaeomagnetic measurements of crustal deformation in New Zealand, I had good reason to believe him. Indeed, I've discovered for myself that an academic move hasn't meant a completely clean slate, because I've brought a fair amount of unfinished business with me to the Southern Hemisphere for my postdoc.

Some of this was unexpected — it turns out that the small matter of 5,000 miles hasn't stopped some students I'd been supervising from soliciting my advice. But I also find my old research still claiming my attention. I have papers on New Zealand tectonics and the growth of magnetic minerals in marine sediments to correct, finish or even start writing, and I still find myself adding new papers on these subjects to my to-read pile.

Having more publications is nice, but this is not my only motivation: I'm also convinced that the best way to develop as a researcher is to move away from a single-minded focus on one project, and keep several lines of inquiry open. This means following up on the questions raised by my past research as well as my current project. Who knows, maybe one day I'll also find myself supervising a PhD... using palaeomagnetism to measure deformation in New Zealand. ■

Chris Rowan is a postdoc in the geology department at the University of Johannesburg, South Africa.

NATURE|Vol 447|31 May 2007

MOVERS

Carla Shatz, director, BioX, Stanford University, Stanford, California

2000–07: Professor and chair, Department of Neurobiology, Harvard Medical School, Boston, Massachusetts
1992–2000: Professor of neurobiology, University of California, Berkeley
1978–91: From assistant to full professor, Stanford University, Stanford, California

Carla Shatz has blazed a trail of firsts for women in neuroscience. Along the way, she has mentored a number of female scientists. She thinks that most girls are interested in science, but many simply lose interest under pressure from society.

After a chemistry degree from the Radcliffe Institute for Advanced Study at Harvard, she received a Marshall Fellowship to University College London and spent the next two years learning about physiology and biological systems.

Shatz returned to the United States to receive the first doctorate in neurobiology from Harvard — in part because she thought the move would offer broader opportunities. Being mentored by David Hubel and Torsten Wiesel, who were conducting Nobel-prizewinning science detailing how the brain is wired, made it an even smarter move.

"I got interested in this incredibly beautiful computational machine, and I wanted to see how the heck it was wired up during development," she says.

For her first job, Shatz's strategy was to go to the place that really wanted her. When Stanford University came calling, Shatz accepted. There, she became the first female basic scientist to be granted tenure at the medical school.

After 13 years at Stanford, she spent the next eight at the University of California, Berkeley, where she tried to establish an interdisciplinary neuroscience institute. Frustrated by a lack of administrative support, Shatz accepted an offer to be the first woman to chair Harvard's neurobiology department. "I couldn't turn it down because I felt I was on a mission to represent women at the highest levels of the university," she says. But she says the competition for space and resources were difficult to handle.

Having always hoped to return to Stanford, Shatz will soon be leading its BioX programme, which merges science, medicine and technology. Stanford neurobiologist William Newsome describes Shatz as an ideal leader for BioX because she has used a host of approaches — from anatomy to molecular biology to physiology — to show that brain development is not solely under genetic control. Shatz says that she sees her biggest challenge as integrating BioX with other departments.

Former colleagues eagerly await her return. "BioX already has fruit, but Carla will grow that tree and expand its influence," says Susan McConnell, a former graduate student of Shatz, adding that one of her many exceptional talents is facilitating long-lasting communities. ■

Virginia Gewin

MENTORS & PROTÉGÉS

Award-winning commitment

Walter Dodds, a biology professor at Kansas State University in Manhattan, likes to do research and to analyse data, but he likes encouraging his students even more. This year, the university's biology graduate student association presented Dodds with the Outstanding Graduate Faculty award. Among the criteria for this honour were teaching skills, mentorship, the cultivation of learning opportunities, participation in graduate student committees and publication and grants record.

Dodds teaches limnology and aquatic and microbial ecology, sciences that arouse his enthusiasm in the classroom and on field trips to nearby streams in the Konza Prairie. His contagious zeal makes learning come naturally.

Dodds says his goals are to "nurture a love for scientific research", and "make students successful in their career paths". Students find this fosters a passionate and professional attitude towards their work. As well as serving on almost every committee of students whose research relates to aquatic ecology, Dodds recently worked with other faculty members to create a course on professional skills in biology — a core requirement for incoming graduate students.

"His door is always open for us," says one of the many students who wrote in to support Dodds's nomination, adding that he "keeps in touch with his past students, meeting up for dinner at conferences".

Other supporters note that Dodds is always eager to work on field trips and lab demonstrations — even in cleaning up — and that he strives to end all his graduate courses with a publication that lists all students as authors. Quick to stem the tide of self-doubt that engulfs all graduate students at some point, he is understanding and flexible towards their family lives. During hours off, he invites students to the pub, or to enjoy his harmonica-playing in the locally renowned Red State Blues Band.

Dodds's professional accolades are impressive. His research has so far secured several million dollars in extramural support for the university, and his long list of publications documents a robust career of quality science. It was not his career, however, that prompted so many letters of support to the award committee. It was his intangible knack of mentorship, the ability to convey to students the idea that, in fact, they are his finest achievement. ■

Ron VanNimwegen is a PhD candidate in biology at Kansas State University.

POSTDOC JOURNAL

Loose ends

My PhD supervisor once told me that your thesis is never really completed. Given that his PhD, like mine, involved palaeomagnetic measurements of crustal deformation in New Zealand, I had good reason to believe him. Indeed, I've discovered for myself that an academic move hasn't meant a completely clean slate, because I've brought a fair amount of unfinished business with me to the Southern Hemisphere for my postdoc.

Some of this was unexpected — it turns out that the small matter of 5,000 miles hasn't stopped some students I'd been supervising from soliciting my advice. But I also find my old research still claiming my attention. I have papers on New Zealand tectonics and the growth of magnetic minerals in marine sediments to correct, finish or even start writing, and I still find myself adding new papers on these subjects to my to-read pile.

Having more publications is nice, but this is not my only motivation: I'm also convinced that the best way to develop as a researcher is to move away from a single-minded focus on one project, and keep several lines of inquiry open. This means following up on the questions raised by my past research as well as my current project. Who knows, maybe one day I'll also find myself supervising a PhD... using palaeomagnetism to measure deformation in New Zealand. ■

Chris Rowan is a postdoc in the geology department at the University of Johannesburg, South Africa.

NATURE|Vol 447|31 May 2007

MOVERS

Carla Shatz, director, BioX, Stanford University, Stanford, California

2000–07: Professor and chair, Department of Neurobiology, Harvard Medical School, Boston, Massachusetts
1992–2000: Professor of neurobiology, University of California, Berkeley
1978–91: From assistant to full professor, Stanford University, Stanford, California

Carla Shatz has blazed a trail of firsts for women in neuroscience. Along the way, she has mentored a number of female scientists. She thinks that most girls are interested in science, but many simply lose interest under pressure from society.

After a chemistry degree from the Radcliffe Institute for Advanced Study at Harvard, she received a Marshall Fellowship to University College London and spent the next two years learning about physiology and biological systems.

Shatz returned to the United States to receive the first doctorate in neurobiology from Harvard — in part because she thought the move would offer broader opportunities. Being mentored by David Hubel and Torsten Wiesel, who were conducting Nobel-prizewinning science detailing how the brain is wired, made it an even smarter move.

"I got interested in this incredibly beautiful computational machine, and I wanted to see how the heck it was wired up during development," she says.

For her first job, Shatz's strategy was to go to the place that really wanted her. When Stanford University came calling, Shatz accepted. There, she became the first female basic scientist to be granted tenure at the medical school.

After 13 years at Stanford, she spent the next eight at the University of California, Berkeley, where she tried to establish an interdisciplinary neuroscience institute. Frustrated by a lack of administrative support, Shatz accepted an offer to be the first woman to chair Harvard's neurobiology department. "I couldn't turn it down because I felt I was on a mission to represent women at the highest levels of the university," she says. But she says the competition for space and resources were difficult to handle.

Having always hoped to return to Stanford, Shatz will soon be leading its BioX programme, which merges science, medicine and technology. Stanford neurobiologist William Newsome describes Shatz as an ideal leader for BioX because she has used a host of approaches — from anatomy to molecular biology to physiology — to show that brain development is not solely under genetic control. Shatz says that she sees her biggest challenge as integrating BioX with other departments.

Former colleagues eagerly await her return. "BioX already has fruit, but Carla will grow that tree and expand its influence," says Susan McConnell, a former graduate student of Shatz, adding that one of her many exceptional talents is facilitating long-lasting communities. ■

Virginia Gewin

MENTORS & PROTÉGÉS

Award-winning commitment

Walter Dodds, a biology professor at Kansas State University in Manhattan, likes to do research and to analyse data, but he likes encouraging his students even more. This year, the university's biology graduate student association presented Dodds with the Outstanding Graduate Faculty award. Among the criteria for this honour were teaching skills, mentorship, the cultivation of learning opportunities, participation in graduate student committees and publication and grants record.

Dodds teaches limnology and aquatic and microbial ecology, sciences that arouse his enthusiasm in the classroom and on field trips to nearby streams in the Konza Prairie. His contagious zeal makes learning come naturally.

Dodds says his goals are to "nurture a love for scientific research", and "make students successful in their career paths". Students find this fosters a passionate and professional attitude towards their work. As well as serving on almost every committee of students whose research relates to aquatic ecology, Dodds recently worked with other faculty members to create a course on professional skills in biology — a core requirement for incoming graduate students.

"His door is always open for us," says one of the many students who wrote in to support Dodds's nomination, adding that he "keeps in touch with his past students, meeting up for dinner at conferences".

Other supporters note that Dodds is always eager to work on field trips and lab demonstrations — even in cleaning up — and that he strives to end all his graduate courses with a publication that lists all students as authors. Quick to stem the tide of self-doubt that engulfs all graduate students at some point, he is understanding and flexible towards their family lives. During hours off, he invites students to the pub, or to enjoy his harmonica-playing in the locally renowned Red State Blues Band.

Dodds's professional accolades are impressive. His research has so far secured several million dollars in extramural support for the university, and his long list of publications documents a robust career of quality science. It was not his career, however, that prompted so many letters of support to the award committee. It was his intangible knack of mentorship, the ability to convey to students the idea that, in fact, they are his finest achievement. ■

Ron VanNimwegen is a PhD candidate in biology at Kansas State University.

POSTDOC JOURNAL

Loose ends

My PhD supervisor once told me that your thesis is never really completed. Given that his PhD, like mine, involved palaeomagnetic measurements of crustal deformation in New Zealand, I had good reason to believe him. Indeed, I've discovered for myself that an academic move hasn't meant a completely clean slate, because I've brought a fair amount of unfinished business with me to the Southern Hemisphere for my postdoc.

Some of this was unexpected — it turns out that the small matter of 5,000 miles hasn't stopped some students I'd been supervising from soliciting my advice. But I also find my old research still claiming my attention. I have papers on New Zealand tectonics and the growth of magnetic minerals in marine sediments to correct, finish or even start writing, and I still find myself adding new papers on these subjects to my to-read pile.

Having more publications is nice, but this is not my only motivation: I'm also convinced that the best way to develop as a researcher is to move away from a single-minded focus on one project, and keep several lines of inquiry open. This means following up on the questions raised by my past research as well as my current project. Who knows, maybe one day I'll also find myself supervising a PhD... using palaeomagnetism to measure deformation in New Zealand. ■

Chris Rowan is a postdoc in the geology department at the University of Johannesburg, South Africa.

The inside track from academia and industry

System-ready scientists

Government and industry must do their bit to ensure that universities provide the workforce they want.

Michael Alvarez

However complex a system might be, there is a reasonably reliable three-step approach by which it can be changed or enhanced: design, development and implementation. Each step has its challenges, and the fact that the process can be depicted simply should not to be taken to mean that it is easy. But the tenets of this framework are sound, and can be applied to management, education and workforce development just as well as to programming and product commercialization. Early on in their careers, scientists would do well to consider how they can fit into and contribute to the enhancement of the labour force or, more specifically, the scientific enterprise.

Viewing people merely as operators that form part of a system may seem dehumanizing, but to an extent this model — which quantifies people in the same way that other natural systems are often quantified — can be both helpful and fruitful. Doctoral training in the sciences helps students to achieve a level of command or mastery within a chosen area of research. It is likely that, sooner or later, such students will be looking for some form of employment. In this way, individuals become part of our broader labour market and contribute to our economic system.

"Global competition for prominence in science and technology is intensifying."

The idea that trainees should be prepared or 'engineered' to fulfil certain functions within society might sound mechanical. Some may even raise questions about the individual's free will or the degree to which government and political structures should be responsible for shaping society. For the sake of clarity, the purpose here is to emphasize the importance of our three broad employment sectors (academia, government and industry) working together to construct a healthy society.

In the United States, the need for reform of the education system, from kindergarten through to graduate school, has been well documented. *Rising above the gathering storm* is a 2005 report by the US National Academies that forms the basis for legislation still being considered in Congress. It highlights the fact that global competition for prominence in science and technology is intensifying, and proposes that the United States will need to do a better job of leveraging its greatest assets (higher education, industrial capital and federal resources) to retain a position of leadership. The committee that authored the report recommends, for example, that the United States strengthen its commitment to basic research that "has the potential to be transformational, to maintain the flow of new ideas that fuel the economy, provide security, and enhance the quality of life". Many fundamental elements are required to produce a healthy economy, and support for innovation is a key ingredient. The education and innovation systems and their inputs need tweaking.

One way that universities can help address the nation's needs and simultaneously retain strength in higher education is to gather information from government and industry about their 'design specifications' for incoming talent. By making use of this such information, universities are better informed and positioned to develop graduate training programmes that produce suitably qualified trainees.

As for then 'implementing' scientists and engineers within the workforce, each company or agency typically has its own methods. What is important is that the dialogue takes place; government and industry must provide academia with input about their needs and the barriers they face. The US Food and Drug Administration, for example, may want to expedite the drug and device approval processes, whereas the companies they regulate might want to identify areas of unmet need or methods by which to enter a particular market. Universities can be instrumental in developing people and technologies that offer solutions.

"Government and industry must provide academia with input about their needs and the barriers they face."

And programmes that raise awareness of these challenges and of the various functions across a company or government organization make trainees more prepared to contribute. These initiatives are valuable to organizations taking on new staff, because they won't need to spend as much time in the early stages of a new recruit's employment getting them up to speed.

In this way, companies that forge links with organizations providing higher education find a more 'system-ready' candidate, with a deeper knowledge of the broader scientific enterprise. Only through open, honest communication can academia, government and industry help design, develop and implement changes that will forge a more efficient science workforce. It is hoped that such changes will lead to a system that gives individuals the best opportunities and societies the best chance to maximize discovery and innovation. ■

Michael Alvarez is director of Stanford University's School of Medicine Career Center, California.